THE UNIVERSITY
of EDINBURGH

Investigation of Transcriptional Regulation of *Foxn1* in Fetal Thymic Epithelial Progenitor Cells

Harsh Jayeshkumar Vaidya

Thesis submitted for the degree of Doctor of Philosophy
The University of Edinburgh
August 2015

Declaration

The work in this thesis is my own, except where otherwise stated.

Harsh Jayeshkumar Vaidya

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Clare Blackburn, for giving me the opportunity to carry out this work and for her constant guidance, intellect and support throughout the last four years. Thank you also for pushing me to deliver my best at the project and during the write up. I would also like to convey heart felt gratitude towards Prof. Ian Chambers, Dr. Simon Tomlinson, and Dr. Andrew Smith, my PhD committee members, for their guidance and support. A huge thank you goes out to all members of the Blackburn group, past and present, for all their helpful discussions, coffee and lunch breaks, and social evenings. A special thank you to Nick for being an intellectual inspiration, amazing colleague, and for patiently answering all my questions. Thanks to Kathy for all the discussions and for ordering what seemed like a million PCR primers over the course of my PhD. Thanks to Svetlana for all the coffee and lunch time conversations. A huge thanks to Alison for teaching me dissections, which formed the basis of the entire project. Thanks to Frank, Karen, and Eilidh for all the social events. A special thank you to Frances and Diana for ensuring the smooth day-to-day running of the lab.

I would also like to thank Fiona Rossi, Claire Cryer, Olivia Rodrigues, and Simon Monard for their expertise in flow cytometric cell sorting. Special thanks to Fiona and Claire for all the cakes that accompanied cell sorting. A huge thank you to all the members of the animal unit, in particular Carol, Stephanie, Laura and John, for taking care of the mice and processing many orders. A special thank you to the past members of the CRM social committee, it was a pleasure organising those amazing events with you guys. A massive thank you to Miguel Foronda for his help and guidance with genome editing using CRISPR-Cas9 system. Thanks also to Pablo, Nick and Nicola for advice on ChIP. Special thanks also to Kousa, Duncan, and Florian for all the help with bioinformatics. A huge thank you to the Principal's Career Development Scheme and Edinburgh Global Overseas Research scheme for funding my PhD, it would not have been possible without their generous contributions.

Finally and most importantly, I would like to thank my family and friends for their invaluable support. To my parents for their continued love and support. Thank you for believing in me and for teaching me that anything can be achieved through determination and effort. A huge thank you for providing me the opportunity to study abroad, it would have been impossible with your help. To my sister, who has been an inspiration throughout my life and always gave many words of encouragement. I am blessed to have you guys!

A special thank you to Laurel and Tyson, for being amazing friends and flatmates. Thank you for the innumerable laughs and fun filled moments, for all the adventures so far and the ones yet to come. We need to make that India trip a reality soon! Thanks also to the noisy members of the Kranc lab, who made the Green area lively and fun and to members of Chambers, Wilson, and Kaji lab for the chat, discussions, and fun moments we shared over the years. Thanks to Martyna, for being a true friend, a really inspirational person and for introducing me to Salsa dancing. Special thanks to Dan, Nic, Kousa, Sabine, Mathias, and Tyson for the parties and Sunday brunches, and for being such amazing adventurous people. Finally, a huge thank you to all my friends at CRM and outside, who made life in Edinburgh beautiful and fun these past several years.

Abstract

The thymus in mice and humans originates from the third pharyngeal pouch endoderm. This process is divided into early *Foxn1*-independent stages and later *Foxn1*-dependent stages. *Foxn1* is indispensible for the differentiation of thymic epithelial progenitor cells (TEPCs) as the development of thymus in *Foxn1* mutant mice is arrested around E12.5. The transcriptional changes associated with the developmental of the thymus are poorly understood. In particular, the transcriptional regulation of *Foxn1* in the developing thymic rudiment has not been definitively identified. Recently, *Pax1*, *Pax9*, *Tbx1*, and *E2F*s have been implicated in transcriptional regulation of *Foxn1*. However, with the exception of *E2Fs*, evidence regarding their direct involvement in regulating *Foxn1* expression is missing. Therefore, the aims of this thesis were to study the transcriptional regulation of *Foxn1* through identification of its regulatory regions and studying the transcriptional changes associated with the developing thymus. These aims were addressed through the use of chromatin-immunoprecipitation technique combined with next-generation sequencing and gene expression analyses of the developing TEPCs. The data presented in this thesis identified H3K4me3 and H3K27ac marked *Foxn1* promoter and five H3K4me1 and H3K27ac marked putative enhancer regions. The combination of gene expression analyses and transcription factor binding sites within the above regions suggested *Ets1*, *Isl1*, *Foxc1*, *Nfia*, *Nfib*, *Srf*, *Foxo1*, *Nfatc2*, *Ing4*, *Foxa2*, *Hes1*, *E2Fs*, and *p53* as candidate transcriptional regulators of *Foxn1*. *Nfatc2* appears also to be a target of *Foxn1* that could play an important role in thymus development by regulating a large set of genes. Comparison of wild type and *Foxn1* null thymus showed that *Foxn1* could act as positive regulator of *Pax1* and negative regulator of *Gata3* and *Eya1*, genes important for third pharyngeal pouch development. The comparison of transcriptome of E10.5 and E11.5 third pharyngeal pouch cells and E12.5 TEPCs showed that genes involved in tissue development are downregulated while those involved in antigen presentation, a process important for thymus function, are upregulated during development. These results also demonstrated a decrease in the activity of transcription factor network involving *Hox* genes and an increase in the activity of a network involving *Nfkb*, *Rela*, and *Irf* genes. Analysis of signalling pathways suggested that the NFκB signalling pathway

could be important for thymus development after E12.5 while TGFβ signalling pathway appeared to be detrimental to *Foxn1* expression in thymic epithelial cells. Together, I identified several transcription factors that could be involved in transcriptional regulation of *Foxn1* in TEPCs, several genes that could be a target of FOXN1, changes in transcription factor network and signalling pathways associated with the developing thymic rudiment.

Lay summary

The maturation of T-cells depends on an extensive crosstalk between the developing thymocytes and a functional thymic stroma, which consists of thymic epithelial cells, mesenchymal cells, dendritic cells, and others cell types. Amongst these, the thymic epithelial cells play a critical role in positive and negative selection of maturing T-lymphocytes ensuring generation of a pool of self-restricted and self-tolerant functional T-cells. The development of the thymus from third pharyngeal pouch depends on the transcription factor, Foxn1. Transcription factors are proteins that can regulate the expression of genes (transcription) within a cell. The transcriptional regulation of *Foxn1* in TECs is currently not well understood. Thus, the aim of this thesis was to use high-throughput and genomic approaches to investigate the transcriptional regulation of *Foxn1* in thymic epithelial progenitor cells, which are capable of giving rise to all TEC subpopulations. Using analysis of fetal wild type and *Foxn1* null thymi, I interrogated the expression pattern of candidate genes for those consistent with candidate transcriptional regulators of *Foxn1*. This led to identification of genes whose expression in TEPCs is influenced by presence or absence of FOXN1 and also of genes whose expression profile matches that expected of a transcriptional regulator of *Foxn1*. I then identified TEPC-specific putative gene regulatory elements using chromatin-immunoprecipitation (ChIP), a technique that identifies proteins bound to DNA, combined with next generation sequencing. ChIP for histone (proteins around which DNA is wrapped in the cell) marks characteristic of active promoter and enhancer regions (genomic regions that regulate gene expression). Four putative enhancer regions and a promoter region were identified for *Foxn1*. Bioinformatics analysis of the *Foxn1* promoter and enhancers identified several transcription factors binding sites within each of these regions. Furthermore, the transcriptional changes associated with developmental progression from third pharyngeal pouch cells to thymic epithelial progenitor cells were investigated using RNA (products of gene expression) sequencing. This allowed identification of genes characteristic of each developmental time point, and genes and signaling pathways regulated differentially during development. The RNA-seq data also suggested that most of the genes with physical binding sites within identified *Foxn1* promoter and putative enhancers are expressed in the developing thymic rudiment. Collectively, the above approaches identified a signaling pathway which, when modulated *in-vitro* or *in-vivo,* can affect the expression of *Foxn1* within thymic epithelial cells, possibly by the binding of its effector molecules to the binding sites identified within *Foxn1* promoter and/or putative enhancers. Finally, I also present a list of transcription factors that represent good candidates for further testing their involvement in regulation of *Foxn1* expression.

Table of contents

Abbreviations

3PP	third pharyngeal pouch
β-gal	β-galactosidase
bp	base pairs
BMP	Bone Morphogenic Protein
cDNA	complementary deoxyribonucleic acid
ChIP	chromatin immunoprecipitation
CMJ	cortico-medullary junction
cTEC	cortical thymic epithelial cell
DC	dendritic cell
DMSO	dimethylsulphoxide
DN	double negative
DP	double positive
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
E	embryonic day
EpCAM	epithelial cell adhesion molecule
ES	embryonic stem
GFP	green fluorescent protein
gRNA	guide RNA
FACS	fluorescence-activated cell sorting
FCS	foetal calf serum
FGF	fibroblast growth factor
FITC	fluorescein isothiocyanate
H&E	Hassal's corpuscle

IL	interleukin
IRF	interferon
IP	intraperitoneally (Chapter 6)
IP	immunoprecipitation (Chapter 4)
IRES	internal ribosome entry site
K	keratin
kb	kilobase pairs
LIF	leukaemia inhibitory factor
MHC	major histocompatibility complex
mRNA	messenger RNA
NC	neural crest
NCC	neural crest cell
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PWM	position weight matrix
QRT-PCR	quantitative real time polymerase chain reaction
rlog	regularized logarithm
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
SD	standard deviation
seq	(next-generation) sequencing
SP	single positive
TEC	thymic epithelial cell
TEPC	thymic epithelial progenitor cell
TF	transcription factor

TGF	transforming growth factor
TN	triple negative
TSS	transcription start site
WT	wild type

1. Introduction

1.1 Regulation of gene transcription in eukaryotes

Regulation of gene expression is fundamental for every biological process. Genes within any cell are constantly being turned on or off depending on numerous cell intrinsic and extrinsic factors. Perturbations in regulation of this process results in abnormal behavior of cells, leading to disorder and diseases. Reflective of its importance to cell survival and function, gene regulatory mechanisms have evolved to become exquisitely sophisticated. François Jacob and Jacques Manod pioneered the first gene regulatory model in the early 1960s. This model laid the foundation for the concept of gene products being involved in regulation of transcription. The current model of gene transcription regulation involves integration of cellular history and extracellular environment occurring at chromatin level and mediated by functionally diversified *cis*-regulatory elements, such as promoter, enhancers, silencers, and insulators. Among these, promoters and enhancers have received great interest from the research community owing to their importance in gene transcription regulation. Promoters are typically defined as DNA region in the immediate vicinity of the transcription start site, which acts as a docking site for pre-initiation complex and other proteins. Enhancers, on the other hand, are DNA elements that could improve the transcription of target genes over huge genomic distances and independently of both orientation and position. Enhancers also govern the spatial and temporal regulation of its target gene(s) expression. Given this property of enhancers, their identification in the past often relied upon discovery of disease-associated mutations occurring at a considerable distance from the candidate gene. However, recent advances in our understanding of epigenetics and its role in regulating gene transcription has allowed identification of combinatorial chromatin signatures that allow prediction of enhancers, and also of promoters. Genome-wide analysis of histone modifications, facilitated by approaches such as chromatin immunoprecipitation combined with microarray or next-generation sequencing, showed that combinations of histone modifications can be used to identify promoters and enhancers (Heintzman et al. 2007; Creyghton et al. 2010; Rada-Iglesias et al. 2011). These studies showed that promoters and enhancers can be identified by

presence of histone modifications H3K4me3 and H3K4me1, respectively. Furthermore, presence of overlapping H3K27ac histone modification in addition to the above can distinguish between “active” and “inactive” promoters and enhancers.

The promoters and enhancers influence the transcription of target genes by binding of transcription factors (TFs) to specific DNA sequences within these elements. The binding of TFs to DNA is influenced by various factors associated with DNA sequence, such as CpG island, DNA methylation, presence of histones and higher-level DNA structures, and histone modifications (reviewed in (Lenhard et al. 2012; Calo & Wysocka 2013)). Furthermore, some transcription factors, such as FOX family members, have the ability to modify chromatin structures to make the underlying DNA sequence more accessible for other factors . The binding of TFs to specific DNA sequences is also influenced by the presence of other TFs and co-factors, providing additional layer of specificity and selectivity (Lelli et al. 2012). Given that enhancers can be located at huge distances from their target genes, the question of how these elements affect transcription has received a lot of attention in the last few years. One such study showed that cohesion complex could play an important role in bringing together enhancers and promoters by a threading mechanism in which Mediator-bound DNA is bought in close proximity of the TSS (Kagey et al. 2010). Recent studies have also suggested coding exons could act as enhancers for nearby genes, controlling their tissue-specific expression (Birnbaum et al. 2012). Furthermore, there is also evidence of bidirectional transcription from some enhancer elements, suggesting that these non-coding RNAs from enhancers could provide another level of complexity in regulation of transcription (Creyghton et al. 2010; Rada-Iglesias et al. 2011). Finally, the relative contribution of enhancers and promoters in regulation of transcription can be different for different gene classes. Developmental genes are typically regulated at both the enhancer and promoter level, tissue-specific genes predominantly depend on *cis*-regulatory modules, and housekeeping genes have a preference for promoter based regulation (Ernst et al. 2011). Thus, gene transcription results from cumulative effect of various inputs of varying degrees of complexity. This thesis addresses the transcriptional

regulation of a key cell type in the thymus, one of the central organs of the immune system.

1.2 Introduction to the thymus

The thymus, a primary lymphoid organ central to immune system function, is a multi-lobed organ located within the mediastinum of the thoracic cavity, in front of the heart and behind the sternum (Figure 1.1A). The thymus is the main site of T-lymphocyte development, the commitment, differentiation, and repertoire selection of T-lymphocytes happens within the thymus. The classical mouse model *nude*, caused by a recessive mutation in the key transcription factor *Foxn1*, results in the absence of a functional thymus (athymia) leading to a lack of mature, functional T-cells in the periphery and thus severe immunodeficiency (Nehls et al. 1994).

1.2.1 Thymus structure and morphology

The thymus is a multi-lobular organ. Each lobe consists of a central medulla, an outer cortex, and a cortico-medullary junction (CMJ) (Figure 1.1B**).** The thymic lobes are surrounded by a layer of connective tissue, known as the outer capsule, and an inner sub-capsular layer composed of simple epithelium (Boyd et al. 1993). In adults, the developing T-lymphocytes account for most of thymic cellularity (~95%) with the rest being composed of the thymic stroma (Blackburn & Manley 2004). The thymic stroma is a highly complex microenvironment consisting of a number of different cell types, including thymic epithelial cells (TECs), dendritic cells, fibroblasts, macrophages, natural killer cells, and B-cells. The TECs, which form the major component of the thymic stroma, can be further sub-divided based on their location within the thymus and the presence of various surface markers. The epithelial cells of the cortex, also known as cortical thymic epithelial cells (cTECs), are identified by the presence of Ly51 or CD205. On the other hand, the medullary epithelial cells (mTECs) are identified by the presence of UEA1. A small proportion of the $UEA1^{+}$ mTECs also express AIRE, and are thought to be post-mitotic. Mature cTECs and mTECs are thought to be distinguished from their immature counterparts

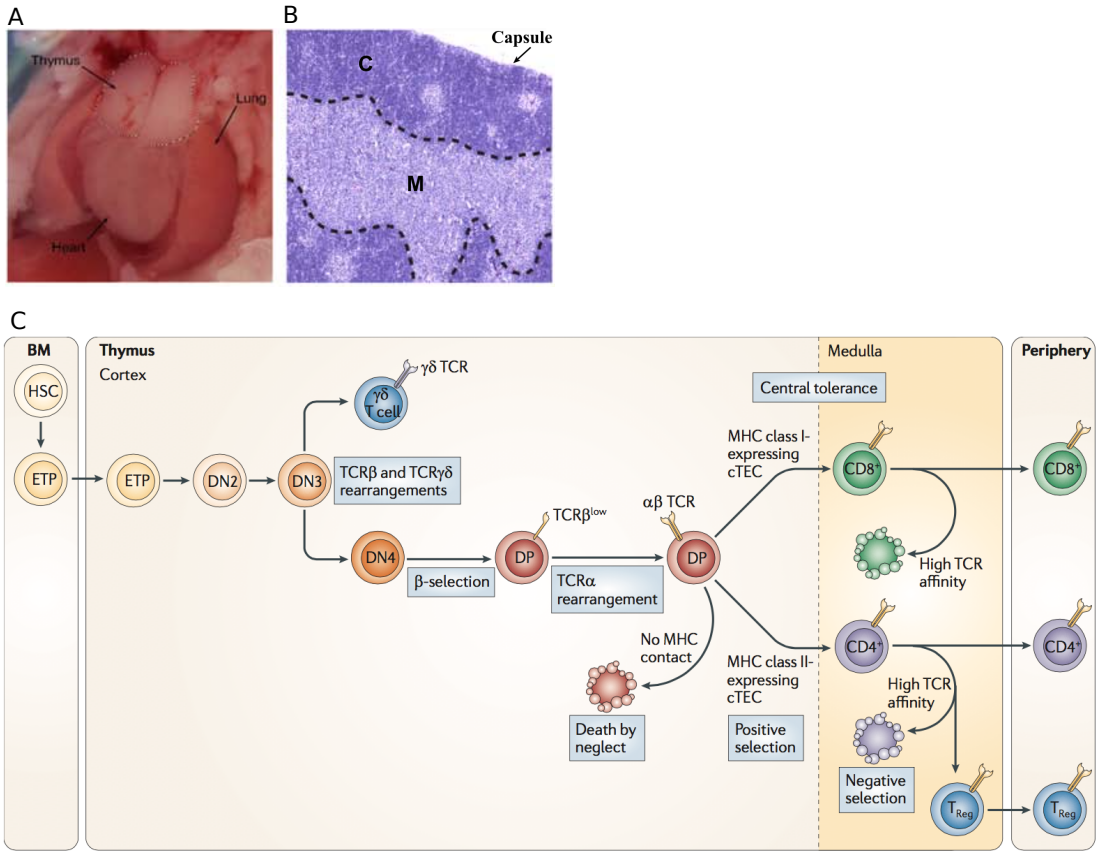


Figure 1.1: Thymus location, structure and the process of T-cell development.
(A) The murine thymus is located within the midline (mediastinum) of the thoracic cavity, in front of the heart and behind the sternum. (B) H&E stain of the murine thymus depicting the darker cortex and lighter medulla. The cortex stains darker due to higher density of thymocytes relative to the medulla. The thymic storma consists of thymic epithelial cells, dendritic cells, macrophages, and fibroblasts. Also shown is the surrounding capsule indicated with an arrow. C: cortex, M: medulla. (A) and (B) adapted from Kelly M, PhD thesis. (C) Schematic of major events in thymocyte development. The early thymic progenitors entering the thymus at the cortico-medullary junction are multipotent. When ETPs lose multipotency, they become double negative (DN) cells, and begin T cell receptor rearrangements. Upon TCRβ gene selection, the DN cells generate double positive (DP) cells, which express a properly rearranged TCR β-chain and both CD4 and CD8 co-receptors. This is followed by TCR α-chain rearrangement and expression of αβTCR. At this stage, a failure to contact MHC molecules presented by cTECs and dendritic cells results in death by neglect. TCRs that bind to MHC Class I molecules lose CD4 expression, wherease those that bind to MHC Class II lose CD8 expression, generating single positive (SP) cells. This process is termed positive selection. If the avidity of binding to MHC-peptide ligands exceeds a certain threshold, the cells are deleted by negative selection in the medulla. Mature SP cells finally migrate out to the periphery via bloodstream. A small percentage of CD4+ thymocytes with an avidity for MHC Class II molecules just below the threshold for negative selection become regulatory T (Treg) cells through upregulation of Foxp3 expression. Adapted from Miller 2011.

by the presence of MHC Class II on the mature cells. A small proportion of the epithelial cells in an adult thymic lobe express PLET1, a surface marker used to identify the common thymic epithelial progenitor cells in the developing mouse thymus. Recent studies in our lab show that the adult PLET1$^+$ cells contain stem cells capable of contributing to TEC networks in a transplantation assay.

1.2.2 The function of the thymus in adaptive immune system

The process of T cell repertoire generation occurs in the thymus, as a series of highly regulated steps. Haematopoietic progenitor cells from the bone marrow enter the thymus via the bloodstream, a process mediated by chemokine signals (Bleul & Boehm 2000; Lind et al. 2001; Liu et al. 2005) (Figure 1.1C). In the adult, these cells are lin$^-$CD44$^+$CD25$^-$Sca-1$^+$c-kit$^+$, and are termed early thymic progenitors (ETPs) (Godfrey et al. 1992; Godfrey et al. 1993; Matsuzaki et al. 1993; Allman et al. 2003). The incoming progenitor population is thought to be heterogenous but recent studies have defined the canonical T cell progenitor cell as the Flt3$^+$ fraction of ETPs, which is highly similar to the LMPP (lymphoid-primed multipotent progenitor) based on transcriptome and potency analysis (Luc et al. 2012). The studies have demonstrated the potential of single ETPs to generate T-cells, B-cells, bone marrow derived natural killer cells, dendritic cells and macrophages (Balciunaite et al. 2005; Porritt et al. 2004; Luc et al. 2012; De Obaldia et al. 2013). The commitment of these progenitor cells to the T-lineage pathway and their subsequent development is a complex process, which depends on their interaction with TECs. The ETPs entering the thymus have a triple negative (TN) phenotype characterized by the absence of the T-cell co-receptors CD4 and CD8, and the CD3 chain of the T-cell receptor (TCR) (Godfrey et al. 1992) (Figure 1.1C). Upon entry at the CMJ, these cells migrate outwards through the cortex towards the subcapsule (Lind et al. 2001). The intrathymic migration of developing thymocytes is directed by chemokines expressed by the thymic stroma (Takahama 2006). During their outward migration through the cortex, these cells are provided necessary signals for their differentiation and proliferation (such as DLL4, IL7), and for the beta and positive selection steps (Lind et al. 2001). The development of these outward migrating TN cells can be sub-

divided into four progressive maturation stages based on the expression of CD44 and CD25. The ETPs entering the thymus is a heterogeneous population capable of generating T-cells, B-cells, macrophages, dendritic cells, and natural killer cells (Allman et al. 2003; Luc et al. 2012). The commitment of ETPs to T-cell fate depends on activation of Notch signalling in these cells through interaction with DLL4, produced mainly by cTECs in the thymus (Koch et al. 2008). The TN cells entering the thymus are $CD44^{+}CD25^{-}$ (TN1) and progress to a $CD44^{+}CD25^{+}$ (TN2) stage during which they undergo proliferative expansion (Godfrey et al. 1993). The transition from TN2 stage to TN3 stage ($CD44^{-}CD25^{+}$) is accompanied by commitment to T-cell fate and T-cell receptor (TCR) gene rearrangement, which generates T-cells of either $\alpha\beta^{+}$ or $\gamma\delta^{+}$ lineage (Godfrey et al. 1993). Upon successful TCRβ rearrangement, the DN3 population migrate to the subcapsular region where they develop into DN4 ($CD44^{-}CD25^{-}$), complete TCR gene rearrangement, undergo further proliferation and subsequently become $CD4^{+}CD8^{+}$ double positive (DP) (Godfrey et al. 1993). The DP cells then migrate back towards the CMJ and undergo 'positive selection', a process mediated by the interaction of DP cells with cTECs. During this selection process, the DP cells are screened for their ability to recognize peptide in the context of self-MHC. A large proportion of the DP cells express αβTCRs that are incapable of peptide:self-MHC recognition and therefore do not receive survival signals from cTECs. Such DP thymocytes undergo apoptosis due to 'death by neglect'. The remaining DP thymocytes, which receive cytokine mediated survival signals, undergo further maturation and ultimately become either CD4 or CD8 single positive (SP) cells depending on the strength and duration of MHC-induced T-cell receptor signalling (Liu & Bosselut 2004). During their migration through the cortex, the maturing thymocytes undergo several rounds of proliferation resulting in a substantial increase in the number of thymocytes within this region. The SP cells then enter the medulla under influence of the cytokines such as CCL21 and CCL19. Within the medulla, the self-reactive thymocytes that express a TCR that bind self peptide in the context of MHC with high affinity are deleted from the repertoire by negative selection, while self-restricted, self-tolerant thymocytes receive further maturation signals. The self-antigen presenting $Aire^{+}$ mTECs and dendritic cells within the thymic medullary compartment play an important role in

this negative selection process. It has been shown that mice carrying mutations in the *Aire* gene suffer from the autoimmune disorder, APECED (autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy). Finally, the mature T-lymphocytes exit the thymus and enter the bloodstream. The pool of naïve T-cells leaving the thymus constitutes a diverse TCR repertoire, enabling these cells to bind peptides from a very broad range of pathogens. The negative selection process ensures that the T-cells entering the peripheral blood are largely tolerant against normal components of the body, which is essential for preventing autoimmune attacks.

1.2.3 Thymus organogenesis

1.2.3.1 Thymic epithelial progenitor cells (TEPCs) in thymus organogenesis

The thymus is a specialized lymphoid organ, which evolved in the vertebrates, except the jawless fish, to support the development of T-cells. The evolution of the thymus coincides with the appearance of VDJ recombination (Boehm & Bleul 2007). The development of the thymus can be viewed as a process of organogenesis and thus can be divided into different stages: positioning, initiation, outgrowth and patterning, and differentiation and migration. Positioning determines the precise location of organ rudiment in the developing embryo; initiation is the visible development of the organ rudiment; outgrowth and patterning involves the generation of regional differences in the growing rudiment and detachment of the organ from its surrounding tissue; and differentiation and migration involves generation of distinct cell types and migration of the organ towards its postnatal position (Manley 2000; Blackburn & Manley 2004). Thymus organogenesis depends on interactions between cells of all three embryonic germ layer origins: endoderm-derived epithelium, neuroectoderm-derived neural crest mesenchyme, and mesoderm-derived hematopoietic cells and endothelial cells of blood vessels.

In mice, the thymus and parathyroid organs develop from a bilateral organ primordia arising from the third pharyngeal pouch endoderm (Blackburn & Manley 2004). These organs develop as an outgrowth (visible from E10.5) from the third pharyngeal pouch resulting from epithelial-mesenchymal interaction between the third pharyngeal pouch endoderm and neural crest derived mesenchyme from the third and the forth pharyngeal arches. Gordon and colleagues (Gordon et al. 2004) showed that the specification of the third pharyngeal pouch endoderm to thymus fate occurs as early as E9.0. The cells of 3PP start to proliferate from E10.0 to form the organ outgrowth, characterised by expansion of both epithelial and mesenchymal compartments. The common thymus/parathyroid primordium begins to separate from the pharynx by E11.5, with the organs subsequently migrating towards their final location in postnatal mice. The thymus and parathyroid domains within this common primordium can be identified by the presence of intra-cellular markers *Foxn1* and *Gcm2* respectively. The expression of *Gcm2* within the third pouch endoderm can be detected from as early as E9.5; whereas, the expression of *Foxn1* is detectable from E11.25 (Gordon et al. 2001). At E11.5, *Gcm2* and *Foxn1* are expressed in a complementary fashion marking the parathyroid- and thymus-specific regions (Gordon et al. 2001). The first ETPs begin to colonize the thymus around E11.5 (Owen & Ritter 1969), under the influence of cytokines secreted from differentiating thymic epithelial cells and mesenchyme (Bleul & Boehm 2000; Wurbel et al. 2001; Liu et al. 2005; Liu et al. 2006).

The embryonic origin of the various sub-populations of the TECs had been a topic of much debate for a number of years. An elegant study in 2004 brought this debate to an end (Gordon et al. 2004). This study, using the classical lineage tracing techniques and grafting of E9.0-E9.5 pharyngeal endoderm under the kidney capsule, proved the endodermal origin of both cTECs and mTECs. The third pharyngeal pouch endoderm gives rise to a common thymus/parathyroid primordium, which separates from the pharynx by E11.5. This process of separation from the pharynx involves apoptotic death of the cells at the junction between the primordium and pharynx (Gordon et al. 2004). By E12.5, the common primordium begins to separate into respective thymus and parathyroid domains, which continue their migration towards

their locations in the adult mouse. At this stage, the thymic rudiment is composed mainly of thymic epithelial progenitor cells (TEPCs), capable of giving rise to all thymic epithelial lineage cells. The existence of a TEPC population was first postulated by Blackburn and colleagues (Blackburn et al. 1996). This study showed that the *nu* (*Foxn1*) gene is required cell-autonomously for the generation of a functional thymic stroma. Moreover, this study also identified that the cells of the nude thymic rudiment had a $MTS20^{+}MTS24^{+}MTS33^{+}$ phenotype, which in normal adult thymus identifies a small population of TECs. Furthermore, this population persisted in nude-wild type chimeric mice and was found to be negative for the expression of markers associated with mature TECs, such as major histocompatibility complex (MHC) class-II (Blackburn et al. 1996). The above discoveries, together, lead to generation of the hypothesis that the developing thymus contained $MTS20^{+}MTS24^{+}MTS33^{+}$ thymic epithelial progenitor cells, which undergo developmental arrest in the absence of *Foxn1*. This hypothesis was later proven when the capability of this $MTS20^{+}24^{+}$ population to form a fully functional thymus upon ectopic transplantation was demonstrated (Bennett et al. 2002; Gill et al. 2002). The antigen bound by mAbs MTS20 and MTS24 was later shown to be PLET1 (placenta-expressed transcript-1), an orphan protein of unknown function (Depreter et al. 2007). Although proving the existence of TEPCs, these studies did not answer the question regarding the presence of a single bipotent progenitor cell population capable of giving rise to both cTECs and mTECs. An alternate possibility to the presence of such a bipotent progenitor is the existence of lineage specific progenitors. Recent studies have demonstrated evidence favouring the existence of a bipotent progenitor (Rossi et al. 2006; Bleul et al. 2006), and more restricted lineage specific progenitors (Bleul et al. 2006). One of these studies transplanted single GFP^{+} $EpCAM^{+}$ cells from E12.5 thymic rudiment, which contain the $PLET1^{+}$ population, into intact, isolated E12.5 thymic rudiments from WT mice and subsequently grafter under the kidney capsule (Rossi et al. 2006). Four out of thirteen grafts exhibited the presence of GFP^{+} mTECs and cTECs four weeks post transplantation, concluding the existence of common bi-potent progenitors. The second study showed that functional thymic tissue containing both cortical and

medullary regions can be generated upon reactivation of a conditional null *Foxn1* allele in single epithelial cells postnatally (Bleul et al. 2006).

Upon separation of the common primordium from the pharynx, it resolves into discrete thymus and parathyroid organs, which continue to develop and migrate toward their final location in the adult body. The first cTECs and mTECs begin to appear in the developing thymic rudiment soon after this and the development of these two compartments proceeds until E15.5 in a lymphocyte independent manner (Bennett et al. 2002; Klug et al. 2002; Shakib et al. 2009). After E15.5, the development of thymic epithelium into an organised 3D structure depends on its interaction with the developing thymocytes, as shown by analysis of *Ikaros*$^{-/-}$ mice, which exhibit absence of thymocytes and a thymic rudiment which is similar in appearance to that at E13.5-E15.5 (Klug et al. 2002). The thymi in these mice is devoid of organised medullary areas defined by $K8^{-}K5^{+}$ TECs, and instead is dominated by more immature $K8^{+}K5^{+}$ TEC clusters. A wild type thymus continues to develop beyond this point with the developmental process being completed around 4 weeks postnatally.

1.2.3.2 Role of neural crest cells (NCC) derived mesenchyme in thymus development

The development of the third pharyngeal pouch and the organs derived from it depends on the interaction between the pouch endoderm and the surrounding mesenchyme. Jiang et al. showed that the mesenchyme that migrates towards and surrounds the third pharyngeal pouch endoderm are of neural crest (NC) origin (Jiang et al. 2000). NC-derived mesenchyme have been implicated in patterning of 3PP endoderm based on the observation that the E12.0-E12.5 *Splotch* mutants, which lack a functional *Pax3* gene and thus functional mesenchyme, have larger thymus domain and a smaller parathyroid domain, as determined by the expression of *Gcm2* and *Foxn1*, compared to WT embryos (Griffith et al. 2009). The thymic lobes in E12.5 *Splotch* mutants were found to have developed normally, with no defect observed in thymocyte migration into thymic rudiment or TEC proliferation or

differentiation, but were ectopically located above the laryngeo-tracheal groove (Griffith et al. 2009). The ectopic location of the thymus in these mutants was likely a result of severe deficiency of NC-derived mesenchymal capsule surrounding the developing 3PP and subsequent thymic rudiment (Griffith et al. 2009), suggesting that the NC-derived mesenchyme are required for separation of the common thymus-parathyroid primordium from the pharynx and/or subsequent migration. The role of NC-derived mesenchyme in migration of thymic rudiment was recently confirmed through the observation that the adult thymus in NC-specific *ephirn-B2*$^{-/-}$ mutant mice but was located ectopically, however no defects were observed in the development of thymus from 3PP, migration of NC-derived mesenchyme, or separation from pharynx (Foster et al. 2010). The NC-derived mesenchyme are found in the outer capsule, inner connective tissue and trabeculae, intrathymic fibroblasts and pericytes, and smooth muscle cells that surround blood vessels, and persist in the adult thymus (Foster et al. 2008; Müller et al. 2008). The normal development of the thymus after E12.5 *in vitro* was arrested in fetal thymus organ cultures stripped of mesenchymal cells (Jenkinson et al. 2003), suggesting an important role for these cells in thymus development. The NC-derived mesenchyme produce FGF7 and FGF10 growth factors that activate the fibroblast growth factor receptor, FgfR2IIIb, which is important for normal proliferation of TECs (Erickson et al. 2002; Jenkinson et al. 2003; Revest et al. 2001). In the absence of functionally competent mesenchyme, as in the case of *Patched* (*ph/ph*) mutants, which lack functional PDGFRα, there is a marked reduction in the expression of Kitl, and Dll4 by the thymic epithelium (Itoi, Tsukamoto, Yoshida, et al. 2007). However, the inference of results from these mutants is complicated as the *Patched* deletion also results in loss of other relevant genes including *c-kit* (Nagle et al. 1994).

More recently, Neves and colleagues showed that culturing quail E2.5 third and fourth pharyngeal pouch endoderm (prior to expression of *Foxn1*) with chick somatopleural mesenchyme (shown to be permissive for development of quail endoderm into thymus and parathyroid) allows initiation of *Foxn1* expression in the endoderm in absence of Noggin but not in its presence (Neves et al. 2012). Although Noggin can inhibit initiation of *Foxn1* expression in the above conditions, there

seems to be no effect of Noggin on *Foxn1* expression from a later stage, E3.0 (~12 hours before detectable *Foxn1* expression in quail embryo), quail endoderm (Neves et al. 2012). Thus, mesenchyme to be necessary for the induction of *Foxn1* in these *in-vitro* culture system. The importance of BMP4 signaling for thymus development in mice has also been a subject of interest for several years. Recently, Gordon et al. used three different Cre-recombinase systems to study the role of Bmp4 in mice. The deletion of *Bmp4* from pharyngeal pouch endoderm and NC- derived mesenchyme using *Foxg1-Cre* (Gordon et al. 2010) does not result in an absence of *Foxn1* expression; however the *Foxg1-Cre* does not give uniform deletion of Bmp4 in the mesenchyme and thus the absence of any phenotype can be due to presence of BMP4 produced by mesenchymal cells failing to undergo recombination. On the other hand, deletion of *Alk3/Bmpr1a* (a *Bmp4* receptor) from TECs resulted in mild thymic hypoplasia suggesting that Bmp signalling may contribute to organ size.

1.2.3.3 Role of vascularisation in thymus development

The first immigration of ETPs into thymic rudiment occurs in the absence of thymic vascularisation, however vascularisation is important for subsequent colonisation events. Thymic vascularisation begins at E13.5 with the migration of endothelial progenitor cells from surrounding blood vessels in to the thymus under the influence of angiogenic stimuli, such as *Vegf* (Leung et al. 1989; Bryson et al. 2013), expressed by TECs and mesenchymal cells (Müller et al. 2005). Subsequently, perivascular cells, such as pericytes and smooth muscle cells, are recruited at E14.5 to the vessel walls via PDGFR-β signalling, and provide structural support to the growing vessel network (Lindahl et al. 1997; Bryson et al. 2013). The NC-derived mesenchyme, surrounding the blood vessels within the thymus (Le Douarin & Jotereau 1975), act as a source of both pericytes and smooth muscle cells (Foster et al. 2008; Müller et al. 2008). Analysis of spatial arrangement of blood vessels, studied using light and electron microscopy, showed that a large artery present at the CMJ branches into arterioles as it penetrates the thymic parenchyma from within the connective tissue trabeculae (Kato 1997). These arterioles in turn connect to a network of capillaries that extend out into the cortex and loop back at the periphery towards the CMJ where

they drain into post capillary venules (Kato 1997). More recently, it has been shown that only small blood vessels are present within the cortex, while large blood vessels are present at the CMJ and within the medulla of an adult thymus (Mori et al. 2007). The large blood vessels at CMJ and within medulla are surrounded by perivascular space, which plays a role in selective transport of haematopoietic precursors and mature thymocytes between thymic parenchyma and blood stream (Mori et al. 2007).

1.3 Molecular regulation of thymus organogenesis

As described above, thymus organogenesis requires interactions between various different cell types and the entire process can be divided into distinct stages, from 3PP formation to generation of functional thymus. Each of these distinct stages is thought to be governed by sets of distinct but overlapping transcription factors which act in a network to support the various organogenetic and morphological changes. Our current understanding of the transcription factors involved at various stages stems mainly from analysis of naturally occurring mutations, gene inactivation studies and *in-situ* hybridisation experiments. These studies have identified several transcription factors, such as *Hoxa3, Pax1*, *Pax9*, *Tbx1*, *Eya1*, *Six1*, *Foxg1* and others, which are important for normal development of 3PP and/or thymus. These factors are discussed in detail below. Wei and Condie recently studied the expression of several transcription factors by performing series of *in-situ* hybridization at various stages of the developing thymic rudiment (Wei & Condie 2011). Their results suggested important roles for *Foxg1* and *Isl1* in thymus development beyond E11.5 and possibly in initiation of *Foxn1* expression. Furthermore, they also showed that the expression of *Gata3*, *Nkx2.5*, and *Nkx2.6* was restricted to thymus domain of the common primordium at E11.25 but was absent at E11.5, suggesting that these genes are important for development of the thymus prior to E11.5 but are not required (or detrimental) for subsequent development. Several signalling pathways have also been implicated in thymus development, including the BMP and FGF signalling pathways. Table 1.1, adapted from Michelle Kelly's thesis, summarizes these factors and our knowledge of their importance in thymus, and this is further discussed in detail below.

Gene	Gene Expression	Mutant Phenotype	Role
Tbx1	Pharyngeal enoderm and mesodermal core of the pharyngeal arches from E7.5-E12.5.	Absence of the third pharyngeal arch and pouch (Jerome and Papaioannou, 2001).	Segmentation and development of the pharyngeal region
Fgf8	Foregut endoderm and non-neural crest mesenchyme from E8.0, and later in the pharyngeal poches.	Absence of the 3PP (Abu-Issa et al, 2002). Increased apoptosis of NCCs (Macatee et al, 2003).	3PP formation and epithelial-mesenchymal interactions.
Hoxa3	3PP endoderm and surrounding NC-derived mesenchyme from E10.5. Later by TECs.	Athymic - 3PP forms but apoptotic from E11.5 (Manley and Capecchi, 1995 and 1998).	Specification and survival of the 3PP
Eya1	3PP endoderm, adjacent NC-derived mesenchyame and 3rd pharyngeal cleft ectoderm between E9.5 and E10.5.	Athymic at E12.5 (Xu *et al*, 2002).	Outgrowth and patterning of 3PP
Six1	3PP endodemr, adjacent NC-derived mesenchyme and 3rd pharygeal cleft enctoderm from E9.5.	Patterning of 3PP initiatied, however apoptosis leads to athymia from E12.5 (Zou *et al*, 2006).	Patterning and maintenance of the 3PP
Bmp4	Ventral and medial portion of the 3PP and NC-derived mesenchymal cells at E10.75. Restricted to the lateral aspect of the thymic anlage by E11.5. Predominantly restricted to the mesenchymal cells that from the thymic capsule from E12.5.	Reduced mesenchymal cell condensation and partial absence of the thymic capsule. Delay in thymus and parathyroid separation and failed or reduced organ migration (Bleul and Boehm, 2005, Gordon *et al*. 2010).	Required for the mesenchymal-TEC cross talk and organ migration
Gcm2	Dorsal portion of the 3PP from E9.5.	Lack parathyroids but display normal thymus development (Gunther *et al*., 2000).	
Foxn1	Ventral portion of the 3PP from E11.25. Later by all TECs.	Athymic and hairless. Thymic primordium forms but arrests between E11.5 and E12.5. No colonisation by lymphocytes (Nehls *et al*. 1994, Nehls *et al*. 1996, Itoi *et al*. 2001).	Regulates TEC differentiation
Pax1	Foregut endoderm from E8.0 and all four pharyngeal pouches, including the 3PP, from E10.5. Restricted to a subpopulation of cTEC in the adult thymus.	Mildly hypoplastic/cystic thymi, delayed T-cell maturation and a 2-5-fold reduction in T-cell number (Wallin *et al.* 1996).	TEC development and function
Pax9	Foregut endoderm from E8.0 and all four pharyngeal pouches, including the 3PP, from E10.5. Expressed throughout the thymic epithelium in the adult, with potentially more expression in the medullary and subcapsular regions.	3PP retarded from E11.5. Ectopic, cystic thymi located in the larynx, which are apoptotic from E14.5. Few T-cells and deregulated NC-derived mesenchyme and no outer capsule (Hetzer-Egger *et al*. 2002).	TEC development and function and T-cell survival

Table 1.1: Key genes important for thymus development.
Shown is a summary of the expression pattern, null phenotype, and potential role of key genes during thymus development. CMJ: corticomedullary junction, TEC: thymic epithelial cells, 3PP: third pharyngeal pouch, NC: neural crest. Table adapted from Kelly M., PhD thesis.

1.3.1 *Tbx1* (T-box 1)

Tbx1 gene encodes a member of the T-box family of transcription factors and is the key gene deleted in the DiGeorge syndrome in humans, which causes defects in development of the craniofacial and pharyngeal regions (Lindsay et al. 1999; Lindsay 2001). During murine pharyngeal development, *Tbx1* is first expressed in mesoderm at E7.5. Its expression in pharyngeal endoderm, ectoderm, and core mesoderm can be detected between E8.5 and E11.5 (Chapman et al. 1996). The expression of *Tbx1* in pharyngeal region appears to be regulated by the sonic hedgehog (Shh) and Wnt signalling pathways (Garg et al. 2001; Huh & Ornitz 2010). In mice, the absence of *Tbx1* expression during development results in abnormal patterning of the first pharyngeal arch, hypoplasia of the second pharyngeal arch and absence of the third, fourth, and sixth pharyngeal arches and pouches (Jerome & Papaioannou 2001). These defects may be partly due to the absence of *Fgf8* and *Fgf10* expression in *Tbx1* null embryos as these fibroblast growth factors play a role in induction or maintenance of proliferation. This is supported by the observation that pharyngeal endoderm has reduced levels of proliferation in the absence of TBX1 at E8.5 (Xu et al. 2005). This study analysed the temporal requirement of *Tbx1* expression in developing murine pharyngeal regions by using a tamoxifen-inducible Cre recombinase to delete exon-5, encoding part of the T-box domain, from a conditional *Tbx1* allele (Xu et al. 2005) and showed that deletion of *Tbx1* at E7.5 or E8.5 resulted in defects in formation of the third pharyngeal pouch and subsequent absence of thymus at E18.5. However, deletion of *Tbx1* at E9.5 or E10.5 resulted in a hypoplastic thymus; whereas its deletion at E11.5 did not have any effect on the thymus. This study, therefore, predicts a critical time window for *Tbx1* expression in thymus development of between E8.5-E9.5, which coincides with 3PP formation. In addition, the contribution of TBX1 expressing cells to the thymus was demonstrated using fate mapping experiments, where tamoxifen-inducible Cre was knocked into the *Tbx1* locus and used to activate β-galactosidase expression, which demonstrated that cell expressing *Tbx1* at E8.5 contributed to the thymus primordium, whereas those expressing *Tbx1* at E9.5/10.5 showed only a small contribution to the thymus (Xu et al. 2005). This suggests that *Tbx1* is crucial for 3PP formation. Furthermore, it also suggests that cells of E9.5/10.5 third pharyngeal pouch endoderm are

heterogeneous for the expression of *Tbx1* possibly because of down-regulation of *Tbx1* expression. This is supported by the observation that *Tbx1* expression is restricted to the parathyroid domain by E10.5 (Vitelli et al. 2002; Dooley et al. 2007). Further evidence for importance of temporal regulation of *Tbx1* expression comes from a study showing that ectopic expression of *Tbx1* in thymus fated cells within 3PP results in down-regulation of *Foxn1* expression and reduced proliferation of these cells leading to reduction in thymus size (Reeh et al. 2014). A consequence of defects in pharyngeal pouch development upon deletion of *Tbx1* at E7.5/8.5 is the absence of *Pax1* expression at E10.5. However, whether this loss of *Pax1* expression is a direct effect of the absence of *Tbx1* expression or an indirect effect resulting from failure of normal pouch development is not clear.

Recently, *Ripply3* was shown to represses the transcriptional activity of *Tbx1* in *in-vitro* Luciferase assays (Okubo et al. 2011). The expression domain of *Pax9* (and also possibly *Pax1*) was increased in the pharyngeal region of E9.5-E10.5 *Ripply3*$^{-/-}$ mice (Okubo et al. 2011). Furthermore, a 3.7kb promoter region of *Pax9* was shown to be TBX1 responsive in Luciferase assay. The results from the above two studies together suggest that the expression of *Tbx1* may be required for the initiation of *Pax9* expression. Furthermore, temporal regulation of *Tbx1* expression is necessary as *Ripply3*$^{-/-}$ embryos have an ectopic and hypoplastic thymus, located in the oropharynx, a phenotype similar to *Pax9* null thymus (Okubo et al. 2011).

1.3.2 *Fgf8* (Fibroblast growth factor 8)

Secreted molecules of the Fibroblast growth factor (Fgf) family of proteins signal via cell surface tyrosine kinase receptors that have both mitogenic and differentiative effects. In the developing pharyngeal regions, *Fgf8* is expressed from E8.0 in foregut endoderm and NC-derived mesenchyme, and later also in pharyngeal pouches (Abu-Issa et al. 2002). A recent study characterized expression domains for various genes involved in *Fgf* signalling using *in-situ* hybridization (Gardiner et al. 2012). This revealed that various *Fgf* genes and downstream targets of FGF signalling are expressed in a highly localized fashion in the E10.5 3PP with *Fgf3*, *8*, and *15* being

expressed in posterior pouch and *Fgf10* being expressed by NC-derived mesenchyme surrounding the pouch (Revest et al. 2001; Gardiner et al. 2012). Homozygous null mutation of *Fgf8* phenocopies the 22q11 deletion syndrome (DiGeorge Syndrome) with respect to defects associated with development of pharyngeal arches and pharyngeal pouches (Frank et al. 2002), however, null mutant embryos die at E8.5 preventing their use in studying the role of FGF8 in thymus development. On the other hand, *Fgf8* hypomorphic embryos show a failure in formation of third and fourth pharyngeal pouches, suggesting a role for FGF8 in 3PP formation (Abu-Issa et al. 2002). The observed defects in pharyngeal pouch formation are consistent with a role for FGF8 as a downstream signalling effector of *Tbx1* (Vitelli et al. 2002; Huh & Ornitz 2010; Guo et al. 2011). Conditional deletion of *Fgf8* in *Hoxa3*-expressing cells, using *Fgf8*$^{loxpP/loxP}$;*Hoxa3*IRESCre, which deletes *Fgf8* from 3PP at E9.5, resulted in hypoplasia and ectopia of the thymus and increased apoptosis of NCCs (Macatee et al. 2003). Thus, *Fgf8* expression in both NC-derive mesenchyme and endodermal cells of 3PP is required for normal pouch formation.

Other *Fgf* genes important for epithelial-mesenchymal interactions during thymus development include *Fgf10*. *Fgf7* and *Fgf10* are both expressed by stromal mesenchyme and its receptor *FgfR2IIIb* is expressed by TECs, suggesting these *Fgfs* could play an important role in interaction between these two cell types. Indeed, deletion of either *Fgf10* or its receptor results in severe thymic hypoplasia and developing thymic epithelium is developmentally arrested at E12.5 (Revest et al. 2001). A role for *Fgf7* in regulating TEC function was supported by a study demonstrating that administration of FGF7 to thymocyte depleted E16.0 thymic lobes in culture resulted in expansion of mTECs and restored normal mTEC associated chemokine expression (Erickson et al. 2002). These studies suggest that *Fgf7* and *Fgf10* may have important roles during later stages of thymus development, possibly promoting proliferation of TECs. However, the role of FGF signalling between E8.5 and E12.5 remained unknown until a recent study studied the effect of FGF signalling using *Spry1* and *Spry2*, two FGF feedback antagonists mutants (Gardiner et al. 2012). This study showed that *Sprouty* genes are expressed at low levels throughout most of the E10.5 3PP and a loss of this expression in *Spry1*$^{-/-}$

Spry2$^{-/-}$ results in increased expression of FGF signalling target genes *Etv4*, *Etv5*, and *Dusp6* across the E10.5 3PP, suggesting increased FGF signalling, and reduced expression of *Gcm2* in the parathyroid domain (Gardiner et al. 2012). *Spry1*$^{-/-}$*Spry2*$^{-/-}$ mutants also show defects in apoptosis of endodermal cells at the junction between the common primordium and the pharynx at E11.75 leading to the failure of the common primordium to completely separate from the pharynx (Gardiner et al. 2012). The loss of *Spry1* and *Spry2* in these mutants results in reduced *Bmp4* expression at E10.5, resulting in a delay in initiation of *Foxn1* expression. However *Sprouty* mutant embryos also show reduced expression of *Fgf10* by mesenchyme surrounding the developing thymus, resulting in a hypoplastic thymus (Gardiner et al. 2012). Thus, dynamic regulation of FGF signalling is required for normal thymus and parathyroid development.

1.3.3 *Hoxa3* (Homeobox protein A3)

Hoxa3 is a member of the homeobox family of transcription factors that are involved in specification of lineage identity (Krumlauf 1994). It is expressed in 3PP endoderm and NC-derived mesenchyme from E9.5. *Hoxa3* is not required for the formation of 3PP (Manley & Capecchi 1995; Manley & Capecchi 1998), however, the expression of *Bmp4* is severely reduced in 3PP endodermal cells of the *Hoxa3*$^{-/-}$ thymi, while the expression domain of *Fgf8* is expanded compared to WT thymi (Chojnowski et al. 2014). Analysis of *Hoxa3*$^{-/-}$ embryos showed that expression of *Pax1* is initiated normally but is not maintained at E11.0, suggesting that *Hoxa3* is not required for initiation of *Pax1* expression but is required for its subsequent maintenance (Manley & Capecchi 1995). The organ specific domains undergo initial pattering in absence of HOXA3 as evident from the presence of *Foxn1* and *Gcm2* expression within the cells of the common primordium, however, the initiation of *Foxn1* expression is delayed whereas *Gcm2* expression is not maintained after E10.5 in *Hoxa3*$^{-/-}$ embryos (Chojnowski et al. 2014). This led to failure of parathyroid cell to undergo differentiation and loss of thymus through apoptosis, resulting in an absence of both parathyroid and thymus in *Hoxa3*$^{-/-}$ mice (Chojnowski et al. 2014). In mice, *Hoxa3*, together with *Hoxb3* and *Hoxd3*, has also been shown to be involved in normal

migration of the common thymus-parathyroid primordium (Manley & Capecchi 1998).

1.3.4 *Pax1* (Paired box protein 1)

Pax1 is a member of the paired box family of transcription factors, whose expression in 3PP endoderm can be detected from E10.5 (Wallin et al. 1996). *Pax1*, and also *Pax9*, are atypical members of the paired box family in vertebrates as they lack the homeodomain (DNA binding domain independent of paired box domain), which is found in all other *Pax* genes. At E12.5, most (if not all) TEPCs express *Pax1*. However, the expression of *Pax1* is restricted to a small subset of cTECs and cells of the subcapsule in adult thymus (Wallin et al. 1996). This was demonstrated through whole mount *in-situ* hybridization and immunohistochemical analysis of embryonic and adult (6-weeks old) thymi. Interestingly, the cTECs that express *Pax1* appear to be immature TECs as determined from an absence of the mature TEC markers MHC Class-II or ER-TR4 (Wallin et al. 1996), suggesting that *Pax1* might be maintained in undifferentiated progenitor cells population. The thymu in $Pax1^{-/-}$ mice is mildly hypoplastic with a 2.5-fold reduction in thymocyte numbers. This defect mainly affects the $CD4^{+}8^{+}$ double positive and $CD4^{+}$ single positive thymocytes (Wallin et al. 1996). The $Pax1^{-/-}$ mouse thymus showed a reduction in both the total number of MHC Class-II^{+} TECs and also the level of expression of this differentiation marker (Su & Manley 2000), which could explain the defects observed in maturing thymocytes. This suggests a role for *Pax1* in the TEC development. Analysis of $Hoxa3^{+/-}$, $Pax1^{+/-}$, $Pax1^{-/-}$, and $Hoxa3^{+/-}Pax1^{-/-}$ mice has shown that the $Hoxa3^{+/-}Pax1^{-/-}$ have a markedly more severe thymus phenotype, including delayed separation of the common primordium from the pharynx, increased thymic hypoplasia due to reduced proliferation and increased apoptosis (Su & Manley 2000). This phenotype in $Hoxa3^{+/-}Pax1^{-/-}$ mice is more severe than the $Pax1^{-/-}$ phenotype, indicating that *Hoxa3* and *Pax1* act in the same genetic network to support the development of the thymus.

1.3.5 *Pax9* (Paired box protein 9)

Pax9 is the most closely related member of the paired box family of proteins to *Pax1*. It is expressed in 3PP from E9.5 (Neubüser et al. 1995; Peters et al. 1998). Compared to *Pax1*$^{-/-}$, the *Pax9*$^{-/-}$ mice exhibit a more severe phenotype with a transient thymic rudiment located ectopically within the larynx, and a complete absence of ultimobranchial bodies. Interestingly, the thymic primordium in *Pax9* null mice develops as an ectopic polyp-like structure in the larynx (Hetzer-Egger et al. 2002). This ectopic thymus is extremely hypoplastic and fails to migrate to upper mediastinum. The expression of *Foxn1* is detectable at E14.5 in *Pax9*$^{-/-}$ mice and the thymic rudiment appears to be colonized by heamatopoetic progenitors. However, the thymic rudiment appears highly disorganised at E16.5, is severely reduced in size and contains large number of apoptotic cells (Hetzer-Egger et al. 2002). Thus, *Pax9* is required for correct positioning of the thymus, and may play a role in survival and development of TECs.

Given that *Pax1* and *Pax9* genes are related by duplication, it is expected that there exists some redundancy between their functions in the thymus. Functional redundancies between these two genes has been demonstrated during vertebral column development in a gene dosage-dependant manner (Peters et al. 1999). In this system, *Pax1* can fully compensate for the absence of *Pax9*, while *Pax9* can partially rescue defects in *Pax1* deficient mice. Recently, studies in our lab have shown that there indeed are functional redundancies between *Pax1* and *Pax9* in thymus (Michelle Kelly, unpublished).

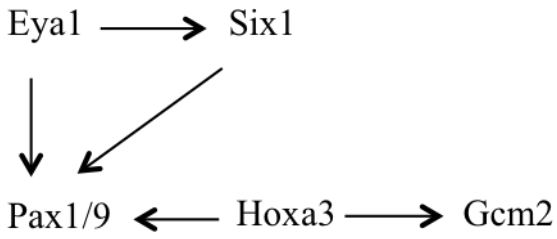
1.3.6 *Eya1* (Eyes Absent 1) and *Six1* (Sine oculis homeobox 1)

Eya1 and *Six1* belong to the *Eyes Absent* and homeobox Six families of proteins respectively. *Eya1* encodes a transcriptional co-activator that is often co-expressed with the SIX1 transcription factor (Xu et al. 1997). *Eya1* is expressed in the second, third, and fourth pharyngeal pouch endoderm, pharyngeal arches (one to four), and the surface ectoderm at E9.5-E10.5 (Kalatzis et al. 1998). *Six1* is expressed in the

pharyngeal endoderm, mesenchyme, and surface ectoderm from E9.5. The expression of *Six1* in E9.5 third pharyngeal pouch endoderm and the surface ectoderm of third clefts is markedly reduced in *Eya1*$^{-/-}$ embryos showing that the expression of *Six1* in these regions is *Eya1*-dependent (Xu et al. 2002).

The common thymus/parathyroid primordium is absent in E12.5 *Eya1*$^{-/-}$ embryos (Xu et al. 2002). Similarly, no common primordium is detectable in E12.5 *Six1*$^{-/-}$ embryos (Zou et al. 2006). However, the common thymus/parathyroid primordium is present at E11.5 in both *Six1*$^{-/-}$ and *Six1*$^{-/-}$*Six4*$^{-/-}$ embryos. The common primordium is smaller in these embryos and consists of more apoptotic cells than their wild-type counterparts, with these defects being more severe in *Six1*$^{-/-}$*Six4*$^{-/-}$ than *Six1*$^{-/-}$ mutants (Zou et al. 2006). Thus, it would appear that the patterning of third pharyngeal pouch endoderm into common thymus/parathyroid primordium is initiated in *Six1*$^{-/-}$ and *Six1*$^{-/-}$*Six4*$^{-/-}$ embryos but undergoes apoptosis by E12.5. Furthermore, the expression of *Gcm2* (at E10.5) and *Foxn1* (at E11.5) was reduced in *Six1*$^{-/-}$ thymi and their expression was absent in *Six1*$^{-/-}$*Six4*$^{-/-}$ embryos suggesting defects in maintenance and differentiation of the cells within the common primordium. Studies on *Hoxa3*, *Pax1/9*, *Eya1*, and *Six1/4* mutant mice indicate a transcriptional network involving these genes. *Hoxa3* expression in third pouch endoderm or NC-derived mesenchyme was unaffected in *Pax1*$^{-/-}$*Pax9*$^{-/-}$ and *Eya1*$^{-/-}$*Six1*$^{-/-}$ embryos; placing *Hoxa3* at the top of this regulatory network. *Eya1* expression was maintained at E10.0 in *Six1*$^{-/-}$*Six4*$^{-/-}$ embryos. Also, the expression of both *Eya1* and *Six1* was found to be normal in *Pax9*$^{-/-}$ and *Pax1*$^{-/-}$*Pax9*$^{-/-}$ mutants. Thus, *Eya1* and *Six1* expression do not require *Pax1/9* and so appear to be above *Pax1/9*, with *Eya1* being above even *Six1* in the transcriptional network. Although *Pax1* expression was present in *Eya1*$^{-/-}$, *Six1*$^{+/-}$ and *Six1*$^{-/-}$ (lower than that in *Six1*$^{+/-}$ genotype) embryos at E9.5/10.5, its expression is absent in *Eya1*$^{-/-}$*Six1*$^{-/-}$ embryos. Similarly, *Pax9* expression was reduced at E10.5 in *Eya1*$^{-/-}$*Six1*$^{-/-}$ embryos, further confirming that *Pax1/9* are downstream of *Eya1* and *Six1* in this transcriptional network (Zou et al. 2006). However, further work is required to elucidate how these transcriptional factors interact and co-operate in the development of the third pharyngeal pouch endoderm and its subsequent organs, the

thymus and parathyroid. The interactions between these transcription factors are summarized in the network diagram below:



1.3.7 BMP (Bone Morphogenic Proteins)

As described above, a short pulse of BMP has been shown to be required for initiation of *Foxn1* expression (Neves et al. 2012), suggesting that BMPs play an important role in development of the thymus. BMPs belong to the TGFβ superfamily of proteins that play important roles during development of various organs. During thymus development, *Bmp4* (and also *Bmp7*) is first expressed at by a small number of NC-derived mesenchymal cells in third pharyngeal arch at E9.5 and is subsequently expressed in endodermal cells in the ventral and medial part of 3PP and surrounding mesenchyme by E10.75. *Bmp4* expression becomes restricted to lateral part of thymic anlage by E11.5 and strong expression is maintained at E12.5 (Bleul & Boehm 2005; Patel et al. 2006). Thus, *Bmp4* may be involved in specification of thymus identity in the ventral part of 3PP. *Bmp4*$^{-/-}$ mice die between E6.5 and E9.5 (Winnier et al. 1995), preventing their use for analysis of its role in thymus development. However, expression of the BMP antagonist *noggin* in TECs via the *Foxn1* promoter resulted in a dysplastic thymus that was drastically reduced in size and exhibited an ectopic location (Bleul & Boehm 2005). This suggests that BMP signaling is required for normal growth, migration, and differentiation of the thymus. Furthermore, the importance of *Bmp4* in epithelial-mesenchymal interaction was demonstrated by conditional deletion of *Bmp4* in pharyngeal endoderm and NC-derived mesenchyme using *Foxg1*-Cre;*Bmp4*^{lacZ} (Gordon et al. 2010). This study showed that deletion of *Bmp4* in this system results in normal patterning of 3PP to thymus but a partial absence of thymic capsule and failure of the thymic primordium to migrate, suggesting that *Bmp4* is not required in specification process.

1.3.8 TGFβ (Transforming Growth Factor – beta)

The TGFβ signalling pathway plays important roles during development, morphogenesis, cell proliferation, differentiation, tissue homeostasis, regeneration, and various diseases (Massagué 2012). In mammals, there are three known isoforms of TGFβ: TGFβ1, TGFβ2, and TGFβ3. These isoforms function through the same receptor signalling pathways upon binding to TGFβ receptors. The TGFβ ligands bind to TGFβRII, which subsequently binds to TGFβRI to form an active complex. This complex phosphorylates receptor-associated SMADs (SMAD2 and SMAD3), which form a hetero- or homo-dimer and subsequently bind to SMAD4 (co-SMAD) forming an active SMAD complex. The active SMAD complex is then nuclearized where is regulates the expression of target genes. The genes targeted by SMAD3 (and possibly other SMADs) are determined by cell-type specific transcription factors that direct the occupancy of SMAD3 to the genome (Mullen et al. 2011). Figure 1.2, adapted from Kubiczkova 2012, shows a schematic representation of the TGFβ pathway

In the developing pharyngeal regions, TGFβ is important for the non-neuronal commitment and survival of migrating NCCs (Wurdak et al. 2005; Wang et al. 2006). The NCC in Wnt1-Cre;*Tgfbr2*$^{loxP/loxP}$ and Wnt1-Cre;*Tgfbr1*$^{loxP/loxP}$ mice migrate normally but fail to contribute to the normal development of pharyngeal regions due to defects in differentiation and increased apoptosis, resulting in formation of hypoplastic thymus that is ectopically located in the neck regions (Wurdak et al. 2005; Wang et al. 2006). The expression of TGFβ signalling ligands by the stromal cells of the thymic cortex, particularly the subcapsule and cTECs, is important for maturation of $CD4^{-}CD8^{-}$ DN thymocytes to DP state (Takahama et al. 1994). In the adult thymus, TGFβ signalling in thymocytes plays an important role in generation and survival of thymic T-regulatory cells (Ouyang et al. 2010; Hauri-Hohl et al. 2014). Furthermore, TGFβ signalling has been implicated in thymus involution, because *Tgfb2* expression in thymus increases with age (Sempowski et al. 2000; Hauri-Hohl et al. 2008). Consistent with this idea, enforced reduction of TGFβ

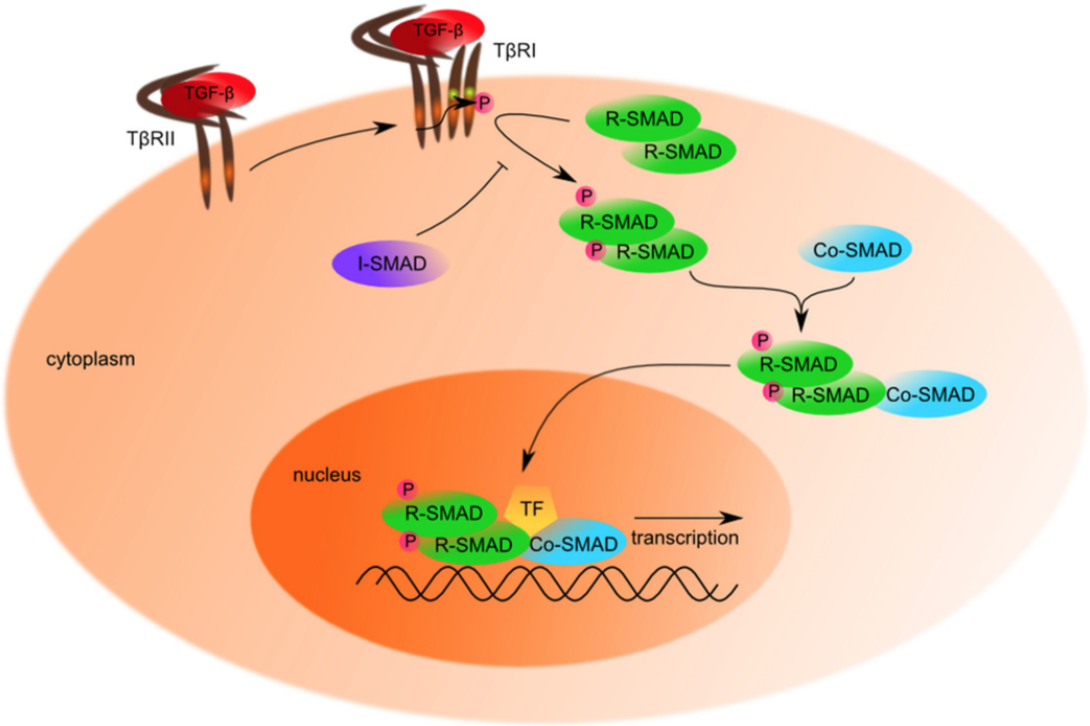
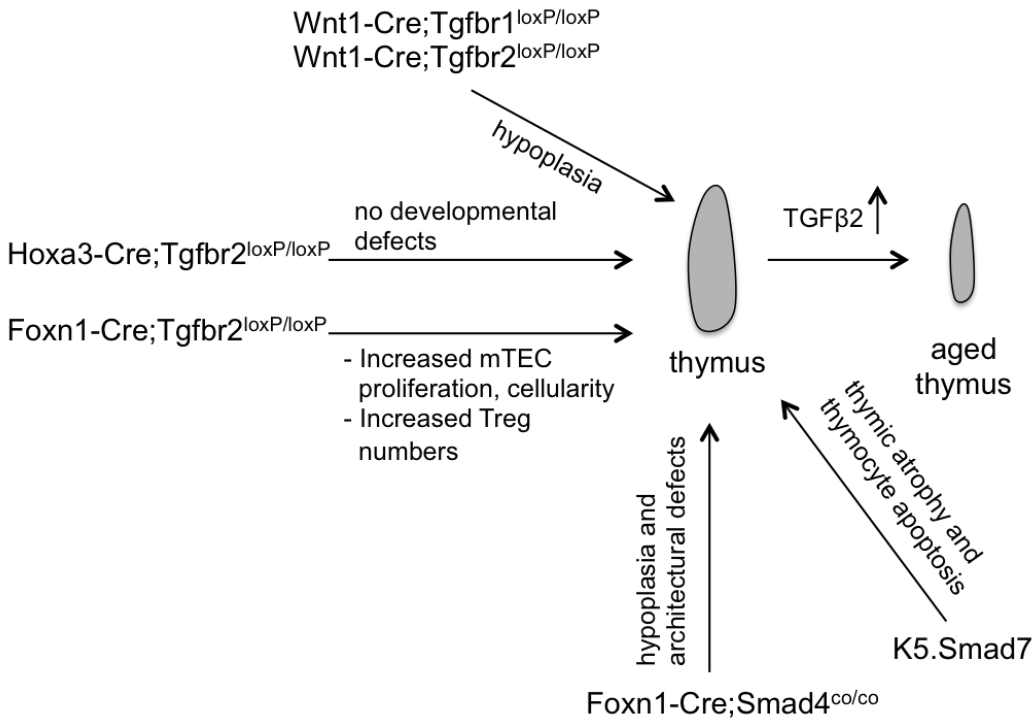


Figure 1.2: Canonical TGFβ signalling pathway.
The binding of TGFβ ligand to its receptor TGFβRII leads it being dimerized with TGFβRI. This receptor dimer complex subsequently phosphorylates intracellular SMAD protiens (SMAD2 or SMAD3), which then form a complex with SMAD4. The SMAD complex is subsequently transported into the nucleus where it binds with specific transcription factors (TF) to regulate the expression of TGFβ target genes. Figure adapted from Kubiczkova et al. 2012

signalling in TECs, HSCs or other non-haemtopoetic cells results in mitigation of thymic involution (Kumar et al. 2006; Hauri-Hohl et al. 2008). Two important studies have shed light on the importance of this pathway in TEC biology. One of these showed that deletion of *Tgfbr2* in *Hoxa3*-expressing cells does not have any adverse effect on thymus development or function, suggesting that TGFβ signalling in 3PP endoderm and NC-derived mesenchyme is dispensable from E10.5 for normal thymus development (Hauri-Hohl et al. 2008). In fact, deletion of TGFβ signalling in TECs in Foxn1-Cre;*Tgfbr2*$^{loxP/loxP}$ mice resulted in an increase in the number and proportion of mTECs due to increased proliferation of these cells (Hauri-Hohl et al. 2014). This increase in mTEC numbers and proportions has a direct functional effect characterised by increased proportion and number of naïve $CD4^+$ and $CD8^+$ SP cells in medulla (Hauri-Hohl et al. 2014).

The TGFβ signalling pathway interacts with various other signalling pathways such as Notch (Andersson et al. 2011), Wnt, Fgf (Paek et al. 2011), and NFκB (Criswell & Arteaga 2007). In primary TEC cultures, deletion of androgen receptor results in decreased *Tgfb1* expression (Lai et al. 2013). TGFβ also negatively affects mTEC differentiation by inhibiting NFκB activation in these cells (Hauri-Hohl et al. 2014). Thus, TGFβ signalling is important for several aspects of thymus biology. However, molecular insights into its functions remain unexplored. In both the classical TGFβ and BMP signalling pathways, signals are transduced via SMAD proteins, which act as transcriptional regulators. Two studies have explored the importance of SMAD proteins in TECs. SMAD4, the SMAD protein common between TGFβ and BMP signalling was studied using *Smad4*$^{co/co}$;Foxn1-Cre mice in which exon-9 of *Smad4* is flanked by LoxP sites (Jeker et al. 2008). Deletion of *Smad4* in these mice resulted in thymic hypoplasia, characterized by a reduction in TEC number and in the number of ETPs within the thymus. While the TEC numbers appeared to recover in mutant mice by 8 weeks old, and no difference is observed in T-cell development, thymus architecture and the ability of thymic stroma to attract ETPs remained compromised throughout life (Jeker et al. 2008). The *Smad4*$^{co/co}$;Foxn1-Cre thymus shows a blurred distinction between cortex and medulla with TECs expressing both K5 and K18, clusters of $ERTR7^+$ fibroblasts, and presence of cysts, a phenotype similar to

that of an involuting thymus (Jeker et al. 2008). This suggests that TGFβ and BMP signalling through SMAD4 play an important role in development and homeostasis of the thymus. Furthermore, the MHC Class-II^{hi} cells in mutant thymus show slightly elevated levels of *Foxn1* mRNA but significantly decreased expression of both *Ccl21* and *Ccl25*, suggesting that there might be some interaction between SMAD4 and FOXN1 proteins as the expression of these cytokine genes is *Foxn1*-dependant (Jeker et al. 2008). The other SMAD protein studied in the context of thymus development is SMAD7, which acts as a negative regulator of TGFβ and BMP signalling pathways. Over-expression of *Smad7* using a transgene driven by the bovine K5 promoter led to severe thymic atrophy and increased thymocyte apoptosis (He et al. 2002). The K5.Smad7 thymi were significantly smaller than controls, and showed extensive thymocyte apoptosis in the cortex, such that the number of $CD4^+CD8^+$ thymocytes was ~50-fold lower in mutant thymi (He et al. 2002). These studies show that mis-regulation of TGFβ and/or BMP signalling in TECs leads to several defects and thus the functions of these pathways in TECs warrants further investigation. The figure below summarizes the defects in the thymi of the above-mentioned TGFβ pathway associated mutants.



1.4 FOXN1 – the master transcription factor in TECs

Foxn1 belongs to the Forkhead family of genes, which encode transcription factors that are essential for normal development, homeostasis, function, and aging of various cell types, organs, and tissues. The evolution and function of Forkhead genes and the importance of *Foxn1* in the thymus are described below.

1.4.1 Evolution of Forkhead genes

The Forkhead gene family is an evolutionarily ancient gene family that derives its name from *D. melanogaster* gene *fork head* (*fkh*). The *fkh* gene was named after the observed phenotype of spiked head appearance in mutant adult flies (Weigel et al. 1989). At the time of its discovery, *fkh* gene was described as encoding a putative transcriptional regulator with a homeotic activity, due to its role in development of terminal segments in fruit fly. The *fkh* gene product showed little homology with any other known class of proteins at the time. Subsequently, the discovery of the rat gene *HNF3* (hepatocyte nuclear factor 3) in 1990, which showed no similarity to any known DNA-binding factors, enabled identification of a highly homologous region between the products of these two genes. This region, consisting of approximately 100 amino acids, was suggested to the DNA binding domain (DBD) of these transcription factors (Carlsson & Mahlapuu 2002). Thus, a new superfamily of transcription factors containing the 'winged helix' or 'forkhead' DBDs (Structural Classification of Proteins (SCOP) classification n°46785), whose members are found in Eubacteria, Archaea, and Eukaryota, was discovered. The term 'winged helix' was coined to reflect the butterfly-like winged structure adopted by DNA-bound Fox proteins. Interestingly, similar wing structures are also observed for DNA interactions with linker histones such as H1 and H5 (Clark et al. 1993). The canonical 'forkhead' domain consists of three N-terminal α-helices (H1, H2, H3), three β-sheets (S1, S2, S3) and two C-terminal 'wing' regions/loops (W1, W2), arranged in the order H1-S1-H2-H3-S2-W1-S3-W2 (Gajiwala & Burley 2000). An additional α-helix is sometimes found in some Forkheads. The DNA binding specificity of the 'forkhead' domain depends on the variable region at the junction of α-helices and wing loops, which interact with bases in minor groove of DNA (Obsil

& Obsilova 2008). To date, over 2000 members of Forkhead superfamily have been identified in 108 species of animals and fungi. The number of Fox genes differs between species presumably due to evolutionary pressures. Among metazoans, 16 Fox genes are found in *C. elegans*, 18 in *D. melanogaster*, 49 in zebrafish, 46 in mouse, and 50 in humans. The nomenclature for these genes was revised in 2000 when a new classification system based on phylogenetic analysis of 172 Fox proteins in 14 species was proposed, in order to unify gene names (Kaestner et al. 2000). The 'forkhead' DBD is highly conserved among the different members of this superfamily and represents the only part of the peptide sequence that can be confidently aligned across all Fox proteins. Earlier phylogenetic trees divided the Fox proteins into 15 classes from FoxA to FoxO based on similarities in FKH domain. These trees were extended over time to include four more classes – from FoxP to FoxS, of which FoxQ, FoxR, and FoxS are vertebrate specific. Figure 1.3 shows the evolutionary tree for Forkhead genes, as adapted from (Hannenhalli & Kaestner 2009). As seen in this tree, the FoxN genes are clustered separately from other classes and the gene most closely associated to *Foxn1* is *Foxn4*, a paralog of *Foxn1*.

Figure 1.4, adapted from Bajoghli et al., shows the evolutionary history of *Foxn1*-like genes (Bajoghli et al. 2009). As seen in this figure, *Foxn4* first appears in cephalochordates (amphioxus), which also contain a more ancient paralog, *Foxn4b*. This ancient paralog, *Foxn4b*, appears to be absent from the genomes of urochordates and all vertebrates. Jawless fish possess a gene very similar to *Foxn4*, termed *Foxn4-like* (*Foxn4L*). *Foxn1* has been proposed to be an ortholog of *Foxn4L* based on protein sequence and short-range synteny relationships. The expression patterns of these genes further support this genealogy. *Foxn4* (i.e. *Foxn4a*) is expressed in the pharyngeal endoderm, among other sites, in amphioxus; *Foxn4L* is expressed in epithelia lining the gill basket in lamprey; and *Foxn1* is expressed in the thymus in cartilaginous fishes and all other jawed vertebrates. However, *Foxn1* has a unique role in the thymus in jawed vertebrate, which cannot be substituted by *Foxn4* (Swann et al. 2014).

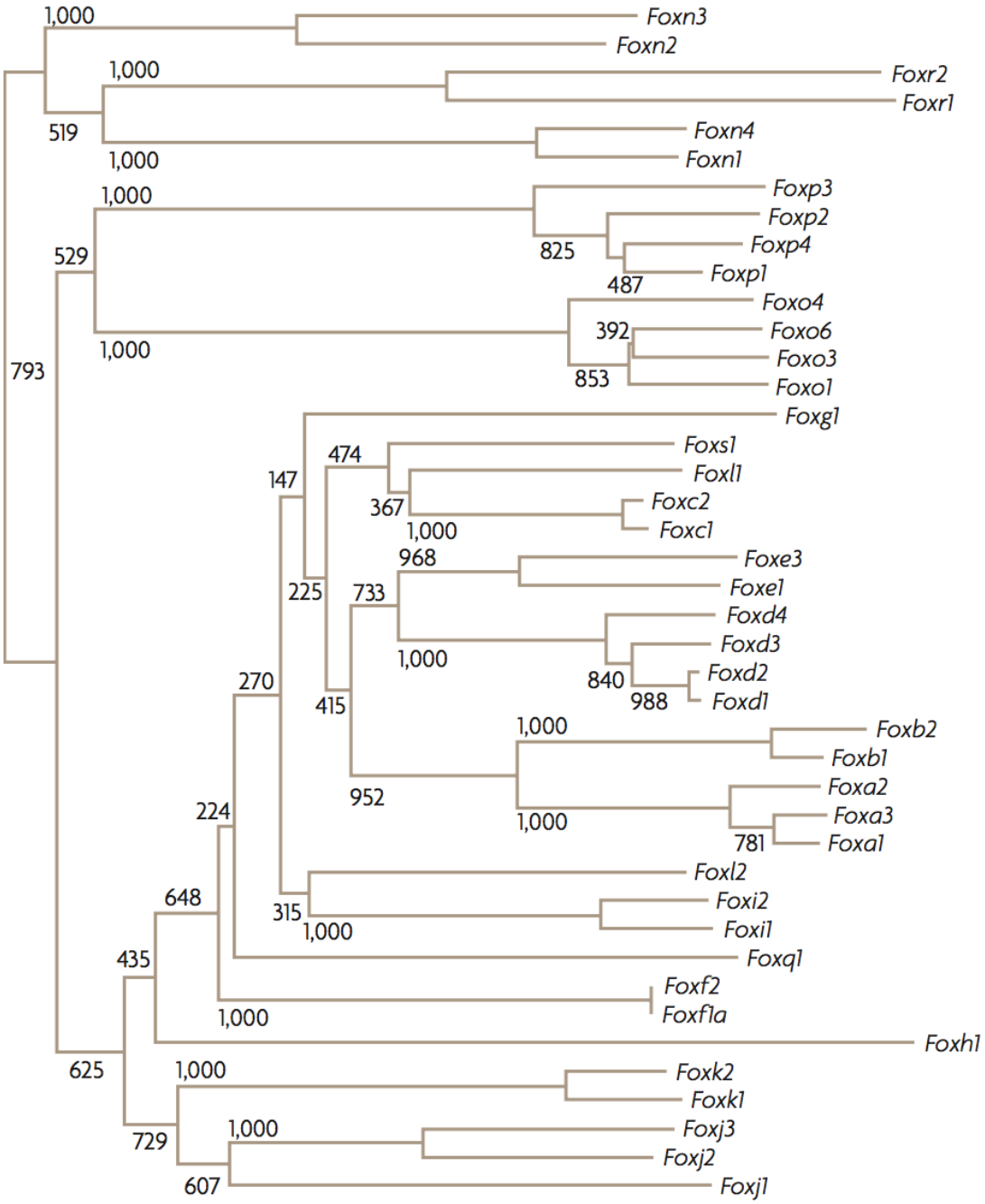


Figure 1.3: Evolutionary tree of mouse forkhead box (Fox) genes.
Shown is a neighbour-joining tree based on multiple alignment of protein sequences of the forkhead domain. Figure adapted from Hannenhalli and Kaestner 2009.

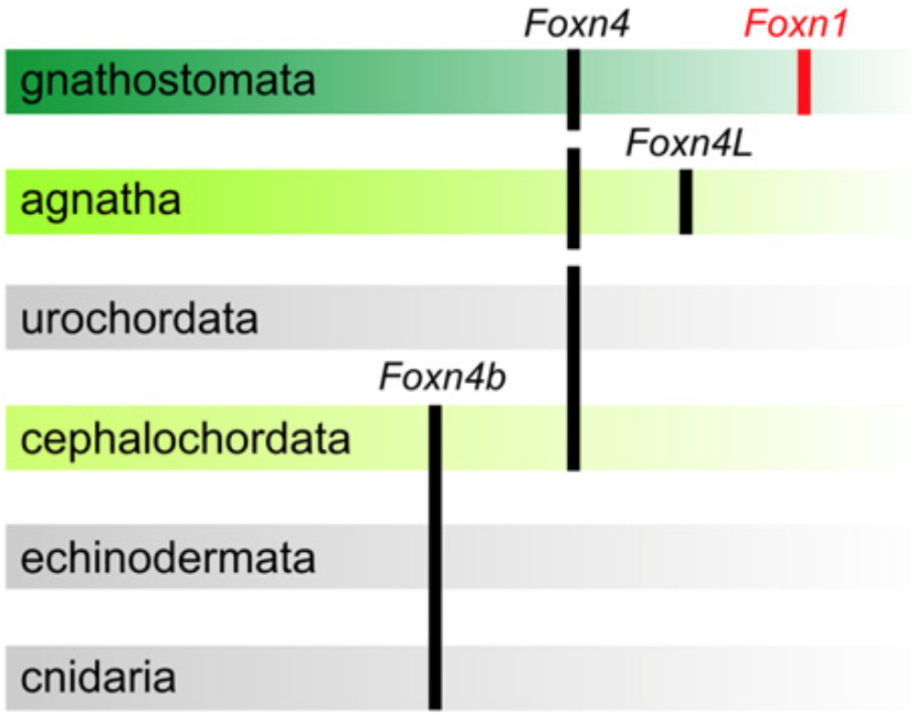


Figure 1.4: Evolutionary history of *Foxn1*-like genes.
The gene *Foxn4b* represents an ancient eumetazoan gene that duplicated in cephalochordata to give rise to *Foxn4a*, which is orthologous to *Foxn4* in urochordates and all vertebrates. *Foxn4L* is a paralog of *Foxn4* in jawless fishes, while *Foxn1* is a paralog of Foxn4 in jawed vertebrates. Figure adapted from Bajoghli et al. 2009.

1.4.2 Functions of Forkhead genes

While the Forkhead proteins share the highly conserved 'forkhead' DBD, the function of these proteins are thought to be determined by the transactivation or repression domains, which show almost no sequence homology between members of this superfamily. Functional diversity is also achieved through differences in interaction partners and in spatio-temporal expression patterns. Thus, while the Forkhead superfamily members have largely distinct functions, some overlap in function can be observed between members of the same sub-group. Mutation of Forkhead genes leads to a range of phenotypes, such as craniopharyngeal defects in *Foxe1* mutants, speech defects in *Foxp2* mutants, a hearing defect in *Foxi1* mutants and others. Several Forkhead genes play an important role in immune system, for example *Foxp3* (development and function of T_{reg} cells), *Foxn1* (TEC differentiation), *Foxj1* (suppression of T-cell activation), and *Foxo3* (regulation of lymphocyte proliferation and apoptosis) (Coffer & Burgering 2004). The FoxO factors are crucial regulators of longevity in *D. melanogaster* and *C. elegans* and strong associations are observed between single nucleotide polymorphisms in the *Foxo1* and *Foxo3a* genes and human longevity (Benayoun et al. 2011).

The Forkhead factors were among the first transcription factors identified as having 'pioneer' function. A transcription factor is considered to be a pioneer factor if it can access its target sequence on nucleosomes and certain forms of compacted chromatin, can bind nucleosome stably and before the binding of other transcription factors or initiation of the target gene expression, and possess chromatin-opening capabilities (Zaret & Carroll 2011). The first forkhead factors identified as having pioneer activity were HNF-3 (FOXA). FoxA factors are required for hepatic specification in mouse; embryos that are deficient for both *Foxa1* and *Foxa2* lack liver bud (Lee et al. 2005). FoxA factors have been shown to be the first TFs to bind the *Albumin* enhancer in gut endodermal cells (Gualdi et al. 1996; Bossard & Zaret 1998). Furthermore, FoxA factors stably bind their target sequences on *in vitro* assembled nucleosome and can open the local nucleosomal domain through the activity of their C-terminal domain (Cirillo et al. 1998; Cirillo et al. 2002). These

elegant studies on the ability of FoxA factors to bind and open nucleosome to regulate target gene expression showed that these TFs could act as pioneer factors. The similarities between the 'winged-helix' structure of the forkhead domain and structures of linker histones is likely why these TFs can displace linker histones from compacted chromatin, even in the absence of SWI-SNF chromatin remodeling complex, and facilitate binding of other transcription factors. Forkhead factors mostly bind to DNA as monomers (Clark et al. 1993), however cases of homodimers (Tsai et al. 2006) and heterodimers (Seoane et al. 2004) have also been documented. Forkhead proteins also interact with non-transcription factor proteins such as co-activators, co-repressors, enzymes and other proteins. Furthermore, some Forkhead proteins are also subject to many posttranslational modifications such as phosphorylation, acetylation, methylation, and ubiquitination (Benayoun et al. 2011). These post-translational modifications affect binding affinity and specificity of their target Forkhead proteins, their nuclear localization and even stability of some of these transcription factors. Finally, Forkhead proteins act as effector molecules for several signaling pathways, converting extra-cellular signals to changes in gene expression (Benayoun et al. 2011).

1.4.3 *Foxn1* in skin

Besides thymus, the only other tissue where the role of *Foxn1* has been studied in any detail is the skin. *Foxn1* expression has been detected in the matrix and the hair shaft of anagen hair follicles (Meier et al. 1999). The skin of *nude* mice show presence of largely normal hair follicles but produce hair shafts that bend and coil, due to abnormal keratinization, and fail to penetrate the epidermis (Mecklenburg et al. 2001). This results in an absence of fur coat in these mice. More recently, *Foxn1* and *Msx2* were shown to be required for *Notch1* expression in the hair follicle matrix, and thus for normal medulla and inner root sheath differentiation (Cai et al. 2009).

1.4.4 *Foxn1* in thymus development

Foxn1 is one of the earliest markers of cells fated towards the TEC lineage within the common thymus/parathyroid primordium. *Foxn1* is expressed at high levels in the ventral aspect of the third pharyngeal pouch endoderm from E11.25, although transcripts can be detected as early as E9.5 by RT-PCR (Nehls et al. 1994; Gordon et al. 2001). Its expression subsequently expands dorsally to cover the entire thymic region of the common primordium. As mentioned above, *Foxn1* was identified as the gene mutated in the classic mouse mutant *nude*, which is characterized by congenital athymia, leading to immune-deficiency, and hairlessness (Flanagan 1966; Pantelouris 1968; Nehls et al. 1994; Nehls et al. 1996). In *nu/nu* mice, a small, cystic, alymphoid thymic rudiment is present in adult mice, and does not support T cell development at any stage in ontogeny. Subsequent to the localization of *nude* locus to chromosome 11 in mice, *Foxn1* (a.k.a. *Whn*) was cloned by position, and it was demonstrated that *Foxn1* transcript in *nu/nu* mutant mice carried a single base pair deletion in the third exon of the gene (which codes for DBD), resulting in the absence of *Foxn1* mRNA presumably due to nonsense-mediated decay. These data suggested that *Foxn1* was the *nude* gene (Nehls et al. 1994). To test this hypothesis, Nehls and colleagues generated a targeted null allele, termed whn^{-}, by inserting a *lacZ-neo* cassette in the third exon of *Foxn1*, close to the site of the spontaneous mutation in *nude* (Nehls et al. 1996). $Whn^{-/-}$ homozygous mice generated using this allele exhibited a phenotype of hairlessness and a thymia, phenocopying the *nude* phenotype, confirming *Foxn1* as the *nude* gene (Nehls et al. 1996). Thus, *Foxn1* became the first member of the forkhead superfamily of genes to be implicated in a specific developmental defect in vertebrates. The targeted disruption study mentioned above also showed that *Foxn1* was not required for formation of 3PP or thymic primordium (Cordier & Heremans 1975; Cordier & Haumont 1980) but was essential for subsequent differentiation of the cells of thymic primoridum into subcapsular, cortical, and medullary epithelial cells (Blackburn et al. 1996). Further insight into the function of *Foxn1* during thymus development came from studies on *nude* mice showing that the absence of functional FOXN1 leads to maturational arrest in TEC at around E12.5 (Blackburn et al. 1996). Furthermore, using *nude* – WT chimeras, this study showed that the *nude*-derived cells persisted in adult thymus but failed to contribute to cTECs or mTECs

and continued to express the antigens bound by MTS-20 and MTS-24 antibodies (later identified as PLET1), leading to the hypothesis that FOXN1 was required for differentiation of TEPCs. Similarly, the skin of chimeric mouse showed contribution from both *nude* and WT cells as determined by presence of patches of skin without hair. Thus, this study showed that *Foxn1* was required cell-autonomously in both skin and thymus.

The common thymus-parathyroid primordium forms normally in *nude* mice and thymus organogenesis is initiated and proceeds normally until E11.5-E12.0 but the thymic primordium fails to differentiate or be colonized by T-cell precursors, which remain in surrounding perithymic mesenchyme (Cordier & Haumont 1980; Itoi et al. 2001). The presence of a thymic rudiment in *nude* and *Whn*$^{-/-}$ mice suggests that the cells within the ventral aspect of the 3PP are specified to a thymus fate well before the onset of high-levels of *Foxn1* expression. Thus, although indispensible for differentiation of TEPCs, *Foxn1* is not required for thymus fate specification of 3PP cells. This is consistent with the observation that transplantation of E9.0 3PP under kidney capsule generates an intact thymus containing both cortical and medullary thymic epithelial compartments (Gordon et al. 2004). Vascularization of developmentally arrested thymic rudiment fails in *nude* mice and these defects result in presence of small, cystic, alympoid thymic rudiment in adult mice (Itoi et al. 2001). It was recently shown that during normal development, endothelial progenitor cells enter the thymus at E13.5, with the supporting mesenchymal cells following at E14.5, and that the timing and extent of migration of these cells is dependent on levels of *Foxn1* expression (Bryson et al. 2013). Interestingly, the expression of *Vegf-a* and *Pdgf-b*, expressed in TECs, vasculature-associated mesenchyme and endothelium in the thymus, is severely reduced under conditions of low *Foxn1* expression, suggesting that the expression of *Foxn1* in TECs is important for normal vascularization of the organ (Bryson et al. 2013). A thymic rudiment with reduced *Foxn1* expression shows fewer capillaries, leaky blood vessels, disrupted endothelium-perivascular cell interactions, endothelial cell vacuolization, and an overall failure of vascular organization at later stages of development. The process of

thymus development can, therefore, be divided into early *Foxn1*-independent (from E9.0 to E12.5) and later *Foxn1*-dependent (after E12.5) processes.

Nowell and colleagues recently studied the requirement of *Foxn1* for various stages of TEC differentiation using a tamoxifen-inducible Cre recombinase to revert a severely hypomorphic *Foxn1* allele back to WT levels (Nowell et al. 2011). The levels of normal *Foxn1* transcript from the hypomorphic $Foxn1^{R}$ allele are around 15% of that from WT allele, allowing generation of an allelic series to study the dose-dependent effects of *Foxn1*. This study showed that increasing levels of *Foxn1* expression are required for TEPCs to progress through multiple intermediate stages of thymic epithelium lineage development in the fetal and adult thymus (Nowell et al. 2011). Reverting the $Foxn1^{R}$ allele at various fetal and adult stages revealed that the *Foxn1* deficient TEPCs can persist in the thymic rudiment until at least 6 months of age in mice and upon reversion are able to generate functional thymic organ with both cTECs and mTECs, showing that these cells are stably locked in a developmentally undifferentiated state under conditions of low *Foxn1* expression (Nowell et al. 2011; Jin et al. 2014). Similar results demonstrating the survival of TEPCs in adult *Foxn1* null mice have been reported using a different revertible *Foxn1* allele termed $Foxn1^{SA2}$ (Bleul et al. 2006). Using a Cre-ERT2 system that exhibits low level activity even in the absence of tamoxifen induction, this study showed that reactivation of *Foxn1* in a single cell in the thymic rudiment of *Foxn1* null mice results in generation of small, well defined, functional medullary areas which are surrounded by cortical spheres (Bleul et al. 2006). This study also demonstrated, using hK14::Cre-ERT2;*Rosa26R*-eYFP reporter mice, the presence of a bipotent progenitor in post-natal thymus, which is capable of giving rise to both cortical and medullary TECs (Bleul et al. 2006). Another hypomorphic allele termed $Foxn1^{\Delta}$, whose transcript lacks an N-terminal domain of FOXN1, is able to support the development of T-lymphocytes but with specific defects at both DN and DP stages characterized by a significant decrease in their proportions and functionality (Su et al. 2003). The $Foxn1^{\Delta/\Delta}$ thymus is severely hypoplastic and cystic and does not contain distinct cortical and medullary regions. This suggests that TEPC differentiation was initiated in $Foxn1^{\Delta/\Delta}$ mice, however cells do not progress through

the entire differentiation process and the developing thymus thus lack functionally differentiated TECs. Indeed, the differentiation of TECs in *Foxn1*$^{\Delta/\Delta}$ appeared to be delayed and arrested as determined by K5, K8, and MTS10 staining (Su et al. 2003). It has recently been shown that TEC differentiation consists of a series of progressive states along the differentiation process, each identified through the expression of cell surface markers (Nowell et al. 2011). The earliest stages of commitment to mTEC cell fate appear to be *Foxn1*-independent, as demonstrated by presence of K5^{hi}Cldn4^{hi} and K5^{-}Cldn4$^{lo/-}$ regions in *Foxn1*$^{-/-}$ thymi (Nowell et al. 2011). Furthermore, mice expressing very low levels of *Foxn1* expression (~15% of WT) showed presence of CD205^{+} cells, suggesting that this level of *Foxn1* expression was sufficient to promoter early stages of cTEC development (Nowell et al. 2011). However stable entry into either differentiation programme and subsequent progression through the differentiation stages depended on increasing dosage of *Foxn1* expression (Nowell et al. 2011). Thus, *Foxn1* appears to be dispensable for earliest fate-choice decisions of TEPCs but is required for subsequent establishment of the differentiation programme.

While the molecular functions of *Foxn1* have not yet been determined in full, it has been shown to be required for expression in TECs of proteins with essential roles in promoting thymocyte development, such as *Dll4*, *Ccl25*, *Cxcl12*, and *Scf* (a.k.a. *Kitl*) (Nowell et al. 2011; Calderón & Boehm 2012). *Ccl25* and *Cxcl12* are chemokines, which are required for attracting ETPs into developing thymic rudiment. *Dll4* is a Notch ligand, which signals to developing thymocyte within the thymus to commit them to T-cell fate and *Scf* is required for thymocytes survival and proliferation. While induced expression of these genes together in *Foxn1*$^{-/-}$ thymic rudiment is able to support, at least to an extent, the generation of CD4^{+} and CD8^{+} SP T-cells, TECs in this transgenic thymic rudiment remain in an undifferentiated state with an absence of medullary regions and very few MHC Class II^{+} cells (Calderón & Boehm 2012). The absence of functional TECs suggest that the SP T-cells observed in this transgenic mice model are likely to have untested functional defects. This indicates that *Foxn1* regulates additional genes that are required for TEC differentiation and function. Indeed, the expression of genes such as *Trp63*, *Pax1*, *Fgfr2IIIb*, *Aire*,

CD40, *CD80*, *CyclinD1*, which have known roles in TEC differentiation or function, and that of several genes involved in Wnt signaling is responsive to levels of FOXN1 in TECs (Nowell et al. 2011; Bredenkamp, Nowell, et al. 2014). Furthermore, FOXN1 may also regulate as yet unknown genes important for thymus development.

A recent study showed that *Foxn1* in mammals has evolved unique functions in TECs that are compensated to some extent by its closest forkhead gene, *Foxn4* (Swann et al. 2014). This study showed that induced expression of *Foxn4*, which is normally only expressed by very few cells in the mouse thymus, in *Foxn1* deficient TECs leads to the development of $CD4^+$ and $CD8^+$ SP thymocytes in the transgenic thymus. However, there is a reduction in the number of DN and SP thymocytes and a proportionate reduction of $CD8^+$ SP thymocytes and a proportionate increase of DN thymocytes (Swann et al. 2014). These transgenic thymi are substantially smaller compared to WT and show restoration of thymic architecture characterized by the presence of cortex and medulla, while also containing several cystic regions (Swann et al. 2014). Both K5 and K8 are widely expressed throughout the adult transgenic thymi, however the expression of Ly51 and UEA1 was restricted to cTECs and mTECs respectively (Swann et al. 2014). The cTEC/mTEC ratio in transgenic thymi is much higher compared to WT resulting from both an increase in the total number of cTECs and a decrease in total number of mTECs. On the other hand, induced *Foxn4* expression was sufficient to rescue the keratinocyte differentiation defect in the skin of *Foxn1*$^{-/-}$ mice (Swann et al. 2014). Contrary to mammalian thymus, the fish thymus expresses similar levels of *Foxn1* and *Foxn4* in TECs. Similar compensation was also observed by the endogenously expressed *foxn4* in *foxn1*$^{-/-}$ medaka thymus (Swann et al. 2014). Interestingly, the induced expression of *Foxn4* in *Foxn1*$^{-/-}$, or a combined induced expression of both *Foxn1* and *Foxn4* in *Foxn1*$^{-/-}$ mouse thymus results in dramatic increase in the number of B cells within the thymus compared to WT thymus (~100 fold in the later transgenic model). B cells are readily detectable in the medaka thymus, which expresses both *foxn1* and *foxn4*. Thus, the specialized function of mammalian thymus to support T-cell development over B-cell development stems from a loss of *Foxn4* expression combined with the maintenance of *Foxn1* expression in TECs.

1.4.5 *Foxn1* in thymus homeostasis and involution

TECs in the adult thymus continue to express *Foxn1* (Nehls et al. 1996), with cTECs expressing higher levels of the gene than mTECs and MHC Class-II^{hi} cells expressing higher levels than MHC Class-II^{lo} TEC in each compartment (Nowell et al. 2011; Ki et al. 2014). Although *Foxn1* continues to be expressed in postnatal TEC, it is not yet clear whether its expression is maintained in all postnatal TECs. The presence of postnatal TECs not expressing *Foxn1* has been suggested from immunohistochemistry analysis of FOXN1, showing an absence of the protein in a proportion of TECs (Itoi, Tsukamoto & Amagai 2007). While the expression of *Foxn1* by all postnatal TECs is still disputed, several studies have demonstrated its importance in maintenance of a functional adult thymus. Interestingly, down-regulation of *Foxn1* expression in the thymic stroma is one of the earliest events in the age-associated degeneration of the thymus (Ortman 2002). This observation suggested that *Foxn1* could play an important role in postnatal thymus homeostasis and function and subsequent thymic involution. This hypothesis was recently tested using a *Foxn1* allele termed *Foxn1*^{lacZ}, which has a IRES-lacZ cassette knocked into the 3` untranslated region of *Foxn1*, that results in normal *Foxn1* expression during fetal and newborn stages but a decline in *Foxn1* expression beginning at about 1 week after birth (Chen et al. 2009). This study showed that, in the postnatal thymus, a reduction in *Foxn1* expression below 50% of WT levels resulted in degeneration of the thymus, characterized by reduced thymus size, deterioration of the cortico-medullary junction, loss of TEC subsets, reduced proliferation of MHC Class-II^{lo} TECs, and defects in thymocyte maturation, in a highly dosage-dependent manner (Chen et al. 2009). The *lacZ/lacZ* thymi demonstrate these adverse effects by as early as 2 weeks of age. The TEC subset most affected by the reduced *Foxn1* expression were the ones with the highest levels of *Foxn1* expression, suggesting that these cells might require high levels of expression of this gene and are sensitive to changes in the level of its expression (Chen et al. 2009). These changes are similar to that observed during age-dependent thymic degeneration, suggesting that reduced *Foxn1* expression in postnatal thymus can accelerate this process. Consistent with this, another study showed that a ubiquitous deletion of *Foxn1* in postnatal mice results in

rapid (within 5 days) thymic atrophy, further supporting the role of *Foxn1* in thymus homeostasis (Cheng et al. 2010). In zebrafish, *mcm2* and *cdca7*, which are down-regulated in presence of *foxn1* morpholinos, appear to be able to partially rescue the T-cell development defect as demonstrated by the expression of *rag1* and *ikaros* in these cells at 4dpf (Ma et al. 2012).

The importance of *Foxn1* in thymus maintenance, as demonstrated by above-mentioned postnatal reduction in *Foxn1* expression studies, is further supported by a study demonstrating the effects of over-expression of *Foxn1* in postnatal thymus. Zook et al. generated a novel *Foxn1* allele, termed *Foxn1*Tg, where a 2.1 kb of mouse *Foxn1* cDNA fragment is expressed under control of the human K14 promoter, resulting in over-expression of *Foxn1* in thymic stroma due to the presence of multiple copies of the transgene (Zook et al. 2011). Transgenic *Foxn1*Tg mice show around 20-fold higher expression of *Foxn1* in TECs than WT mice and the level of expression does not change with age (Zook et al. 2011). Interestingly, expression of the transgene in these mice results in a 7.2-fold higher expression of the endogenous *Foxn1* in 2-months old *Foxn1*Tg mice compared to WT, suggesting a possible auto-regulation of *Foxn1* in TECs (Zook et al. 2011). Trangenic thymi from old mice are considerably bigger in size and show little adipose tissue deposition and a well defined architecture with proper cortical-medullary demarcation (Zook et al. 2011). The thymi from 3-months old *Foxn1*Tg mice showed a 2.5 fold increase in the number of $EpCAM^{+}$ MHC $Class\text{-}II^{+}$ $Ly51^{-}$ TECs and a further 5-fold increase in MHC $Class\text{-}II^{hi}$ subpopulation within these TECs, partly resulting from increased proliferation of TECs in transgenic mice (Zook et al. 2011). Thus, overexpression of *Foxn1* is able to prevent the age-related decrease in TEC numbers and organization. This alleviated involution phenotype also resulted in increased thymic output, as demonstrated by higher number of total thymocytes and higher naïve T-cell output from transgenic thymi as compared to that from WT (Zook et al. 2011). The authors of this study also observed an increase in the number of ETPs in *Foxn1*Tg thymi (Zook et al. 2011), which has been suggested to be a result of increased expression of *Foxn1* in novel population of $Lin^{neg/low}$ $CD45^{+}$ $EpCAM^{+}$ $Sca1^{+}$ $CD117^{-}$ $CD138^{-}$ $MHCII^{-}$ bone marrow cells which in turn is thought

to affect the number of hematopoetic stem cells and multipotent progenitors (Zook et al. 2013). This later study suggests that *Foxn1* affects thymus homeostasis and involution not only through its function in TECs but also through its role in bone marrow.

1.4.6 *Foxn1* as master transcriptional regulator of TECs

The above-mentioned studies showed that *Foxn1* plays an important role in maintenance of transcriptional programme prevalent in TECs, and that decreasing levels of *Foxn1* expression are directly related to age-related changes in this programme. On the other hand, as described in detail below, studies in our lab have shown that *Foxn1* is also able to restore in aged TECs a transcriptional programme similar to that in a young TECs and that *Foxn1* is furthermore able to establish this transcriptional programme in an unrelated cell type resulting in its reprogramming. Bredenkamp and colleagues recently addressed the possibility of regeneration of an aged thymus through overexpression of *Foxn1* by using a novel transgene R26*Foxn1ER*^{T2} encoding a tamoxifen-inducible form of FOXN1 expressed under the control of CAG promoter, from the *Rosa26* locus, upon excision of a stop cassette (Bredenkamp, Nowell, et al. 2014). The expression of this transgene was restricted to TECs in the thymus by using a Cre-recombinase, which is expressed from endogenous *Foxn1* locus. The presence of IRES-*Gfp* following *Foxn1ER*^{T2} in the transgene showed that most TECs in fetal and aged transgenic mice were GFP^{+}, suggesting that most TECs in adult thymus express *Foxn1* (Bredenkamp, Nowell, et al. 2014). Using this transgenic mice model, Bredenkamp et al. showed that tamoxifen treatment of 12- and 24- month old mice for 1 month resulted in *in-vivo* regeneration of thymus, characterized by an increase in thymus size and TEC numbers, total thymocyte numbers (including ETPs), restoration of thymic architecture to that similar in young mice, increased mTEC:cTEC ratio, and increased naïve T-cell output (Bredenkamp, Nowell, et al. 2014). Induced expression of *Foxn1* in this mouse model lead to increased expression of various genes important for TEC biology and function, including *Dll4*, *Ccl25*, *Kitl*, *Pax1*, *Trp63*, *Fgfr2IIIb*, and *Aire*, to levels observed in young thymus (Bredenkamp, Nowell, et al.

2014). This indicates that *Foxn1* is able to drive and restore the transcriptional network prevalent in young TECs. A unique advantage of this approach for thymus regeneration is restoration of architecture and function of old thymus to that observed in young thymus, which has not been observed with other approaches for provoking thymus regeneration such as sex-steroid ablation (Bredenkamp, Nowell, et al. 2014). Furthermore, consistent with previous reports, this study also showed increased proliferation in various TEC subsets as a result of induced *Foxn1* expression, supporting the role of *Foxn1* in TEC proliferation (possibly by regulating *Trp63*, *Fgfr2IIIb*, *Ccnd1* – all of which are involved in cell proliferation) as well as differentiation (Bredenkamp, Nowell, et al. 2014). Despite the role of *Foxn1* in TEPC differentiation, the *Foxn1*Tg and R26*Foxn1ER*^{T2} thymi do not show uncontrolled differentiation of TEC progenitor/stem cells suggesting that other factors regulate the balance between proliferation and differentiation. However, the identity of such factors remains unknown.

In an approach similar to that mentioned above, a full-length *Foxn1* cDNA under the control of CAG promoter was expressed from *Rosa26* locus upon excision of a STOP cassette and the CRE-recombinase is provided by *CreER*^{T2} knocked into the second *Rosa26* locus, allowing excision of the cassette in MEFs (Bredenkamp, Ulyanchenko, et al. 2014). Using this system, Bredenkamp and colleagues showed that by 10 days after initiation of *Foxn1* expression in MEFs, the cells had undergone changes in morphology, adopting a broad polygonal shape characteristic of epithelial cells, suggesting that these cells were being reprogrammed (Bredenkamp, Ulyanchenko, et al. 2014). These FOXN1-induced MEFs were shown to express genes important in TECs and provide a permissive environment for maturation of ETPs to DP and SP thymoctyes *in-vitro*, with similar kinetics to that observed for the well established OP9-DL1 stromal cell line (Bredenkamp, Ulyanchenko, et al. 2014), leading to the conclusion that enforced expression of FOXN1 resulted in induction of TEC identity in MEFs, generating induced or iTEC. The authors further showed that these FOXN1 reprogrammed iTECs were able to generate a fully functional thymus organ, with characteristic thymus architecture, upon transplantation under the kidney capsule together with supporting stromal cells and immature thymocytes

(Bredenkamp, Ulyanchenko, et al. 2014). The reprogrammed iTECs showed induction of endogenous *Foxn1* expression, consistent with the autoregulation of *Foxn1* observation in *Foxn1*Tg mice (Bredenkamp, Ulyanchenko, et al. 2014). While reprogramming of cell fate using a single transcription factor has been demonstrated for other systems, the above study is the first example of generation of a functional organ using such reprogrammed cells. Together, the above two studies suggest that FOXN1 is a master regulator of TEC differentiation, capable of establishing and maintaining a transcription factor network characteristic of young thymus.

1.4.7 Regulation of *Foxn1* expression in TECs

Given the importance of *Foxn1* in the thymus, it is no surprise that several studies have analyzed the regulation of its expression in this organ. As mentioned above, evidence exists for autoregulation of *Foxn1* in TECs, however whether this is a direct or indirect effect of induced *Foxn1* expression remains to be determined. Recently, E2Fs have been shown to be able to bind to their consensus binding site in *Foxn1* promoter *in-vitro* and increased activity of E2F3 *in vivo* was shown to correlate with increased expression of *Foxn1* in TECs (Garfin et al. 2013). Furthermore, genetic analysis has shown that reduction in *Foxn1* expression is sufficient to reverse the phenotype of enlarged thymus associated with *Rb* mutants, further supporting the idea of regulation of *Foxn1* expression by E2Fs (Garfin et al. 2013). Another gene suggested to be involved in regulating the expression of *Foxn1* in the thymus is *Tbx1* (Reeh et al. 2014). Induced expression of *Tbx1* in *Foxn1* expressing cells of E11.5 3PP results in down-regulation of *Foxn1* expression in these cells, suggesting that *Tbx1* could act as a repressor of *Foxn1* transcription in TECs (Reeh et al. 2014). As expected, the reduced *Foxn1* expression in $Foxn1^{Cre};R26^{iTbx1/+}$ thymi leads to dedifferentiation of TECs (loss of *Cldn4* expression at E14.5) resulting in an accumulation of $PLET1^{+}$ cells in the fetal thymus (Reeh et al. 2014). Finally, *Hoxc13* has been suggested to be a transcriptional regulator of *Foxn1* in the hair follicle and skin (Potter et al. 2011).

The regulatory regions governing the expression of *Foxn1* in TEPCs or TECs still remain to be identified. Several studies have investigated the size of the minimal genomic region surrounding the *Foxn1* gene on chromosome-11 that can reproduce the wild-type *Foxn1* expression pattern in skin and thymus. The largest region tested was a 110 kilobase pairs (kb) region containing the entire *Foxn1* locus plus 74kb of 5`-flanking sequence and 12kb of 3`-flanking sequence; this region rescued the *nude* phenotype *in vivo*, suggesting that it contains all regulatory elements required for normal expression of *Foxn1* (Cunliffe et al. 2002). Another study used a cosmid-derived transgene containing 26kb of genomic DNA encompassing the coding exons of *Foxn1* plus 8.5kb of 5`-flanking sequence and 3kb of 3`-flanking sequence (Kurooka et al. 1996). However, this 26kb transgene resulted in rescue of only the hairless phenotype but could not rescue the athymic phenotype of *nude* mice, showing that it lacked at least some of the regulatory regions required for *Foxn1* expression in TECs (Kurooka et al. 1996). As mentioned above, several studies expressing various genes under the control of *Foxn1* promoter element have been published. These studies employ as the *Foxn1* promoter a 30 kb fragment containing the entire upstream sequence between the first coding exon of *Foxn1* (exon-2) and the upstream gene *Slc13a2* (Schlake 2005). This 30kb promoter fragment has been shown to be able to recapitulate *Foxn1* expression pattern in the developing thymus, however it does not reproduce the normal *Foxn1* expression pattern in postnatal and adult thymus (NR Manley, personal communication). Thus, it appears that the regulatory regions governing the transcription of *Foxn1* in fetal TEPCs and TECs are present within the 30 kb region identified by Schlake, while the postnatal expression of *Foxn1* requires additional regions present in the 110 kb region used by Cunliffe. However, the promoter and enhancers governing the expression of *Foxn1* in TEPCs and TECs remain to be definitely identified and similarly the identity of the tissue-restricted transcription factors important for its expression remains elusive.

Studies focusing on improving our understanding of thymus development have so far identified several genes important for this process, however information on molecular insight of this process remains scarce. Particularly, the regulation of *Foxn1* expression in the developing thymus still remains unknown. Analysis of mice

carrying mutations in genes considered to be potential upstream regulators have provided some clues as to potential regulators of *Foxn1*, but with the exception of E2F, evidence of direct binding to *Foxn1* regulatory elements is lacking:

i) Ectopic *Tbx1* expression in *Foxn1* expressing TECs results in downregulation of *Foxn1* expression. However, whether this is a direct or indirect effect remains to be determined (Reeh et al. 2014).

ii) *Hoxa3*$^{-/-}$ and *Eya1*$^{-/-}$ embryos demonstrate a block in primordium formation prior to onset of *Foxn1* expression, hindering analysis of their potential role in regulation of *Foxn1* expression (Manley & Capecchi 1998; Zou et al. 2006). It is now clear that *Foxn1* expression is delayed but initiated in *Hoxa3*$^{-/-}$ thymi (Chojnowski et al. 2014).

iii) *Pax1*$^{-/-}$*Pax9*$^{-/-}$ thymic rudiment does not express *Foxn1* (Blackburn lab, unpublished). However, direct interaction of the PAX proteins with the *Foxn1* locus has not been demonstrated.

iv) Mice mutant for *Rb* family genes show increased *Foxn1* expression in thymus at 3-months of age and E2Fs can bind to their consensus binding sites in *Foxn1* promoter *in-vitro* (Garfin et al. 2013). However, the fetal *Rb* mutant thymi are indistinguishable from controls, suggesting that *Foxn1* may not be regulated by E2Fs during fetal thymus development.

1.5 Aim

The aims of this thesis were therefor to identify candidate transcriptional regulators of *Foxn1* in fetal thymic epithelial progenitor cells, through analysis of the regulatory regions governing *Foxn1* expression in fetal thymic epithelial progenitor cells, in combination with transcriptome and candidate gene analysis.

2. Materials and Methods

2.1 Mice

All animals were housed in Animal Facility at either the Roger Land Building or the Scottish Centre for Regenerative Medicine Building at the University of Edinburgh and treated in accordance with the Animal (Scientific Procedures) Act 1986. Mice were housed in a stabilised environment with a 12-hour light/dark cycle with food and water provided as required.

2.1.1 Mice Mating Set-up and Embryo Collection

The mice embryos used in this study were of C57BL/6 X CBA genotype. Briefly, C57BL/6 females were mated with CBA males by caging together overnight. The females were examined for the presence of a vaginal plug the following morning, which was taken as embryonic day 0.5 (E0.5). Pregnant females were sacrificed at the desired stage and the uterus isolated in an ice-cold tube. The uterus was subsequently dissected into ice-cold phosphate buffered saline (PBS) to release the embryos. All extra-embryonic tissue was discarded at this stage. The embryos were sacrificed by removal of heads.

The *Foxn1* null embryos (termed *nude*) were obtained by intercrossing *Foxn1* heterozygous mice, $Foxn1^{lacZ/+}$. The resulting embryos, a combination of wild type (WT) and transgenic, were collected as described above.

2.1.2 Embryo Developmental Stage Identification

In most cases, the developmental stage of embryos was determined by counting the number of days post vaginal plug examination. More precise classification of the developmental stage, was carried by counting the number of somites in accordance with the guidelines in Karl Theiler's "The House Mouse: atlas of embryonic development" book (Theiler 1989).

2.1.3 Genotyping

The genotypes of embryos resulting from *Foxn1*$^{lacZ/+}$ X *Foxn1*$^{lacZ/+}$ cross were determined by collecting yolk sacs and tail clips for each embryo and adding to a tube containing 50µl tissue and cell lysis buffer. Samples were incubated at 55°C overnight in a shaking waterbath. These samples were then incubated at 95°C for 10 minutes to inactivate Proteinase K and then placed on ice for 5 minutes. Samples were then centrifuged at 13,000rpm for 10 minutes at room temperature and the supernatant, containing genomic DNA (gDNA), was transferred to a clean microcentrifuge tube. The isolated gDNA was used for PCR analysis with allele specific primer pairs.

2.1.3.1 Tissue and Cell Lysis Buffer

10mM Tris-HCL, ph 8.3 (Roche)

50mM Potassium Chloride (KCl)

2.5mM Magnesium Chloride ($MgCl_2$)

0.1mg/ml Gelatin

0.45% NP40

0.45% Tween20

To 50ml in distilled water (Autoclave without NP40 and Tween20)

50µg/ml Proteinase K (Promega) added as required

2.1.3.2 Genotyping PCR Reaction Mix

2µl yolk sac or tail clip gDNA

3µl 10X PCR Reaction Buffer (Qiagen) (containing 1.5mM $MgCl_2$)

0.5µl 100µM dNTPs (Invitrogen)

1µl of 10µM each primer

2.5U Taq DNA polymerase (Qiagen)

To 50µl with sterile water

The following PCR conditions were used: denaturation at 94°C for 5 minutes followed by 30-35 cycles of amplification, 94°C for 30 seconds, 55°C (59°C for exon-3 PCR) for 30 seconds, and 72°C for 30 seconds. Following the completion of amplification cycles, the PCR products were incubated at 72°C for 10 minutes to ensure complete elongation of all PCR products. Finally, the samples were cooled and held at 4°C until analysis by gel electrophoresis.

2.1.3.3 Agarose gel electrophoresis

For gel electrophoresis, 1% agarose gel was prepared by dissolving 1 gram of agarose in 100ml of TBE buffer and 0.5µg/ml ethidium bromide was added to facilitate visualization of DNA. PCR products were mixed with 1X Orange G loading buffer (NEB) before loading in the gel. The agarose gel containing PCR products was typically run at 100-120V in TBE buffer. 0.5µg of 1kb ladder (Invitrogen) or 100bp ladder (NEB) was used as molecular weight marker. The gel was analyzed under a UV transilluminator at 312nm (Gene Flash, Syngene, UK) and photographed either digitally or printed.

2.1.4 Third Pharyngeal Pouch (3PP) and Embryonic Thymus Dissociation

3PP from mice embryos ranging from E9.5 to E10.5 and the thymic primordium from embryos ranging from E11.0 to E12.5 were microdissected under a dissection microscope. Dissected tissue was pooled (where appropriate) in a clean microcentrifuge tube containing ice-cold 1ml of PBS with 5% fetal calf serum (FCS), together termed FACS wash. Collected tissue was then washed twice with PBS containing magnesium (Mg) and calcium (Ca) (Sigma) to remove any traces of FCS, which can inhibit subsequent dissociation steps. Washed tissue was then re-

suspended in 1ml of PBS (+Mg+Ca) containing 0.05mg/ml DNaseI (Roche), 0.7mg/ml collagenase D (Roche), and 1.4mg/ml hyaluronidase (Sigma) enzymes. The samples were then incubated at 37°C for ~15 minutes and mixed by pipetting every 3-4 minutes to aid dissociation of tissue to a single cell suspension. At the end of this incubation time, ~900µl of the suspension was transferred to a clean polypropylene tube (BD Falcon) used for fluorescent activated cell sorting (FACS). The remaining 100µl of the above suspension, which contains smaller, un-dissociated pieces of tissue, was then mixed with 500µl 250µg/ml trypsin containing 0.05mg/ml DNaseI and further incubated at 37°C for 5 minutes, with repeated mixing every minute by pipetting. The resulting single cell suspension was pooled together with that in the FACS tube. The presence of DNaseI in this dissociation protocol ensures that the DNA released from any dying cell is rapidly digested without adversely affecting the dissociation process. The resulting single cells were subsequently washed twice with FACS wash before staining for cell surface markers. All centrifugation steps were carried out at 213 rcf in a refrigerated centrifuge at 4°C.

2.2 Flow Cytometry

2.2.1 Sample Preparation for FACS

A single cell suspension was prepared from tissue samples as described in section-2.1.4 for sorting cell types of interest. All the steps mentioned below were carried out on ice and using ice-cold reagents. All centrifugation steps were carried out at 213 rcf in a refrigerated centrifuge at 4°C. After washing the dissociation enzyme free single cell suspension twice with FACS wash, ~10% of the sample was transferred to a clean FACS tube and used as unstained control for adjusting voltages for lasers being used. For experiments where cells needed to be stained for PLET1, a marker of thymic epithelial progenitor cells (TEPCs) (Depreter et al. 2007), the cell suspension was centrifuged after removal of unstained samples. The supernatant was then discarded and the cell pellet was re-suspended in 200µl of either unconjugated MTS20 or 1D4 antibody, which was generated in the lab, and incubated for 15

minutes on ice. The sample was then diluted with 800µl of FACS wash, centrifuged and the supernatant, containing unbound antibodies, was discarded without disturbing the cell pellets. All subsequent incubations were carried out in dark. The cell pellet was then re-suspended in 200µl of FACS wash containing appropriate concentration of secondary antibody and incubated for 15 minutes on ice. Following this, sample was washed with FACS wash as described above to remove unbound secondary antibodies. Given that the secondary antibodies used were usually always generated against rat (e.g. goat α-rat), the cell pellet obtained after discarding the supernatant from the last step was re-suspended in 200µl of FACS wash containing 5% rat serum to prevent cross-reactivity with any of the antibodies used in subsequent steps. The sample was incubated for 5 minutes on ice and then washed with FACS wash. The resulting cell pellet was then stained for Epithelial Cell Adhesion Molecule (EpCAM; a marker of epithelial cells), CD45 (a marker of haematopoietic cells, including thymocytes), CD31 (a marker of vascular endothelial cells), and Ter119 (a red blood cells marker) in a single step. To this end, the cell pellet was re-suspended in FACS wash containing appropriate concentrations of each of these antibodies and incubated for 15 minutes on ice. Following this, the sample was washed twice with FACS wash and the cell pellet was finally re-suspended in 200-300µl of FACS wash containing appropriate concentration of DAPI (dead cell marker). The cells were kept on ice in dark until sorting.

Single stain controls were used for each fluorochrome for adjusting the compensation between different excitation-emission spectrums. UltraComp beads (BD Falcon) or in-house produced beads were used for single stain controls. A separate bead sample was prepared for each fluorophore used in the experiment. Briefly, one drop of bead was mixed with 1µl of appropriate antibody and incubated on ice for 15 minutes. The bead sample was then washed with FACS wash and then re-suspended in 200µl of FACS wash and kept on ice until analysis. Unconjugated beads were added to stained bead sample as appropriate. FMO controls (full minus one; lacking one colour from the staining panel) were prepared by using 10% of cell suspension transferred into a fresh FACS tube before beginning the staining procedure. The FMO control was subsequently stained for all colours except one,

usually the colour used for identification of PLET1, and used for setting sorting gates.

2.2.2 Cell Sorting

Cell sorting was kindly performed by the FACS facility staff, Simon Monard, Olivia Rodrigues, Claire Cryer, and Fiona Rossi, on FACS Aria-II (BD) using 100µm nozzle. Depending on experimental needs, the cells were either sorted into sterile, DNase and RNase free 1.5µl eppendorf tubes or into 200µl PCR tubes.

For subsequent analysis using CellsDirect One-Step qRT-PCR kit (Invitrogen), 200 cells were sorted into a 200µl PCR tube containing 10µl of reaction mix (see Section 2.3.3). The tubes were immediately centrifuged after sorting and placed on dry ice, to prevent any RNA degradation, to be subsequently transferred to -20°C freezer.

2.2.3 Antibodies

Primary antibodies used for FACS are described in Table 2.2, along with their clone, source, isotype and dilution factor. Secondary antibodies used were raised in goat and conjugated to either alexa fluor-647 or alexa fluor-568, as described in Table 2.3.

2.3 Molecular Techniques: Real Time PCR

2.3.1 RNA Isolation

Before RNA isolation, the bench surface and pipettes were wiped clean with 70% ethanol and RNase OUT (Ambion) to remove any traces of RNase. RNA isolation was carried out using sterile, DNase and RNase-free filtered pipette tips (Alpha Laboratories) and 1.5ml Biosphere SafeSeal Tubes (Sarstedt). RNA from samples that were not limited in cell number (i.e. when not using CellsDirect One-Step qRT-PCR kit) was isolated using RNeasy Mini Kit (Qiagen) according to the

manufacturer's instructions. Isolated RNA was kept on ice for immediate use or transferred to -80°C for long-term storage.

2.3.2 Reverse Transcription and cDNA synthesis

Full-length cDNA was synthesized from total RNA using Superscript-II First-Strand Synthesis kit (Invitrogen), according to the manufacturer's guidelines. First strand synthesis was carried out using oligo(dT) as primers. The reverse transcriptase enzyme was substituted with water to obtain –RT (negative) controls, for each sample, for subsequent PCR reactions.

2.3.3 Protocol For Samples with Low-Cell Numbers

Samples were designated as being low-cell number if it was predicted that sufficient numbers of cells could not be sorted to be able to isolate sufficient quantity of good quality RNA with above-mentioned protocol. Such samples included single embryo sorting for transgenic crosses, cells sorted after cell-culture experiments, and others. Furthermore, samples were treated as low-cell number samples if the sorted cells were to be analyzed using the Fluidigm system. CellsDirect One-Step qRT-PCR kit (Invitrogen) was used to carry out one-step reverse transcription and cDNA synthesis for cells sorted from such samples to avoid material wastage and increase the number of genes whose expression could be tested from a single sample.

Up to 200 cells from low-cell number sample were sorted into a 200µl PCR tube containing 10µl of reaction mix (see below). Care was taken to ensure that cells were sorted in to the center of the tube to maximize the chance of sorting them into the reaction mix, rather than onto the side of the tube. Upon sorting, the PCR tube was immediately centrifuged briefly, to collect all the material (cell droplets and reaction mix) to the bottom of the tube, and then transferred to dry ice for quick freezing of the sample. Samples were stored at -20°C until the pre-amplification step (see next section).

2.3.3.1 CellsDirect Reaction Mix for Cell Sorting

For 200µl PCR tube contains the following:

10µl CellsDirect 2X Reaction Mix (from CellsDirect One-Step qRT-PCR kit, Invitrogen)

0.2µl SUPERase-In (Applied Biosystems)

Up to 200 cells were be sorted in to this reaction mix.

2.3.3.2 Pre-amplification of low-cell number samples for subsequent qPCR analysis

The CellsDirect One-Step qRT-PCR kit has been optimized for gene expression analysis using very few cells, including single cells. The protocol used here is a modification of the protocol for pre-amplification of single cell for use with the Fluidigm platform, based on BioMark User Bulletin 5 (Fluidigm PN 68000107 Rev. A). The protocol was modified for analysis of up to 200 cells, as per recommendations from the Quake lab at Stanford University, USA. PCR tubes containing cells in reaction mix were thawed on ice before use. Following this, 10µl of 4X Assay Mix was added to each tube and pre-amplification was carried out using a PCR machine in accordance with the guidelines in BioMark protocol. Pre-amplified samples were either used immediately or stored at -20°C for later use.

Pre-amplification reaction mix:

To each sample, add the following

5µl 4X Assay Mix

1µl SuperScript III RT/Platinum Taq mix (CellsDirect kit)

4µl TE buffer

-RT (negative) control samples were prepared by substituting the SuperScript III RT/Platinum Taq mix in the above pre-amplification reaction mix with Taq polymerase.

2.3.3.3 100X Assay Mix

100X Assay Mix was made for each gene using forward and reverse primer pairs. 100X Assay Mix can be stored at -20°C for later use.

Each 100µl 100X Assay Mix contains:

20µl 100µM Forward primer

20µl 100µM Reverse primer

60µl sterile water

2.3.3.4 4X Assay Mix

4X Assay Mix was made using 100X Assay Mix for each gene whose expression was to be analyzed for a given sample. The 4X Assay Mix contains equimolar concentrations of primer pairs for all the genes to be analyzed. Of note, the samples could only be pre-amplified once, thus the expression of only the genes included in the 4X Assay Mix could subsequently be analyzed by qPCR. The number of genes that were included in 4X Assay Mix depended on the type of platform used for downstream qPCR analysis. For analysis using LightCycler 480 (Roche), a maximum of 16 genes was analyzed from a pre-amplified sample, due to volumetric constraints. On the other hand, up to 48 or 96 genes could be included in 4X Assay Mix if subsequent qPCR analysis was to be carried out using 48.48 or 96.96 Dynamic Array from Fluidigm, respectively. The 4X Assay Mix could be stored at 4°C for a limited amount of time.

4X Assay Mix contains:

1µl of each 100X Assay Mix

to 100µl using sterile water

2.3.3.5 Pre-amplification Thermal Cycling Conditions

Temperature	Time (min)	Function
50°C	15:00	Reverse Transcription
95°C	02:00	RT enzyme inactivation; Platinum Taq activation
16 cycles of:		
95°C	00:15	Denaturation
60°C	04:00	Annealing/Extension
4°C	Until further use	Cool down

2.3.4 Quantitative PCR using Univeral Probe Library (UPL) on LightCycler 480-II (Roche)

The following protocol was used for gene expression analysis of both non pre-amplified and pre-amplified samples on LightCycler 480-II instrument. 8µl of reaction mix was mixed with 2µl of diluted cDNA in one well of a multiwell qPCR plate. The reaction mix was kept on ice and shielded form light to prevent probe degradation. PCR was performed as per manufacturer's guidelines.

Reaction Mix:

5.0µl 2X Master Mix (Roche)

0.1µl UPL probe (Roche)

0.5µl 10µM Forward primer

0.5µl 10µM Reverse primer

to 8µl with PCR grade water (Roche)

2.3.4.1 Cycling Conditions on LightCycler 480-II

Temperature	Time (min)	Function
95°C	05:00	Denaturation and Hot-start
45 cycles of:		
95°C	00:10	Denaturation
61°C	00:15	Annealing/Extension
4°C	Continuous	Cool down

2.3.5 Quantitative PCR using Univeral Probe Library (UPL) on Fluidigm (BioMark)

The following protocol was used for gene expression analysis of pre-amplified samples on Fluidigm instrument. 10X Assay mix, containing 2µM of each primer and 1µM of the probe, was prepared as described below for each assay. 5µl of each 10X Assay mix was dispensed into individual assay inlets on a primed 48.48 Dynamic Array Chip. 5µl of sample mixes (prepared as described below) were dispensed into individual sample inlets on the same Dynamic Array. The Dynamic Array was loaded into the BioMark instrument and PCR performed as per manufacturer's recommendations.

10X Assay Mix:

2.5µl DA Assay Loading Reagent

0.5µl 100X Primer Pair Mix

0.5µl UPL Probe

1.5µl Water

Sample Mix:

3.64µl TaqMan® Universal PCR Master Mix

0.36µl GE Sample Loading Reagent

2.5µl pre-amplified sample

2.3.5.1 Cycling Conditions on Fluidigm BioMark

Temperature	Time (min)
50°C	02:00
95°C	10:00
40 cycles of:	
95°C	00:15
70°C	00:05
60°C	01:00
4°C	Continuous

2.4 Molecular Techniques: Chromatin Immunoprecipitation combined with next generation sequencing (ChIP-seq)

2.4.1 Sample fixation

Low cell number ChIP-seq was performed using 125,000 cells per ChIP reaction. The number of cells used per ChIP was chosen based on the low-cell ChIP-seq protocol kindly provided by Wysocka Lab at Stanford University, USA.

125,000 cells were sorted into clean, DNase and RNase-free DNA-low Bind 1.5ml Eppendorf tube containing 200µl of FACS wash. After sorting, cells were washed twice with PBS to remove traces of FCS. The cells were then re-suspended in 1ml of sterile PBS (Sigma) and fixed by adding formaldehyde to a final concentration of 1% and incubating at room temperature (RT) for 10 minutes. Excess formaldehyde was then quenched by adding glycine to a final concentration of 0.125M and incubating at RT for 5 minutes. The cells were then washed with ice-cold 1ml of HBSS media containing 1X protease inhibitor cocktail (Roche). The supernatant was discarded and the cell pellet stored at -80°C until further use. All centrifugation steps were carried out at 239 rcf for 5 minutes at 4°C.

2.4.2 Cell lysis and Sonication

All the steps were carried out using sterile filtered pipette tips and on ice unless stated otherwise. Formaldehyde fixed, frozen cell pellets were thawed on ice for 30 minutes. Following this, a three-step gentle lysis protocol was followed to isolate nuclei for sonication. Briefly, the cell pellet was first re-suspended in 500µl of ice-cold LB1 containing 1X protease inhibitor cocktail and incubated at 4°C for 10 minutes on a vertical spinning wheel. The lysis reaction was then centrifuged for 5 minutes at 1350g at 4°C and the supernatant discarded. The pellet was then re-suspended in 500µl of ice-cold LB2 containing 1X protease inhibitor cocktail and incubated at RT for 10 minutes on a vertical spinning wheel. The sample was then centrifuged at 1350g at 4°C for 5 minutes and the supernatant discarded. The pellet containing nuclei was finally re-suspended in 100µl of ice-cold LB3 containing 1X protease inhibitor cocktail and used immediately for sonication.

Sonication was carried out using chilled (ice-cold) BioRuptor Sonicator from Diagenode. Four rounds of sonication was carried out for each sample. The BioRuptor setting for each sonication round was: 10 cycles of sonication at high power, “on” interval = 30 seconds, “off” interval = 30 seconds. The ice-cold water in sonication bath was changed after every sonication round to keep the apparatus

chilled and to avoid heat mediated degradation of sonicated chromatin. The sonicated chromatin was used immediately for ChIP.

2.4.2.1 Lysis Buffer-1 (LB1)

		Final concentration
2.5ml	1M Hepes-KOH pH7.5	50mM Hepes
1.4ml	5M NaCl	140mM NaCl
100µl	0.5M EDTA	1mM EDTA
10ml	50% glycerol	10% glycerol
2.5ml	10% NP-40 (or substitute)	0.5% NP-40
1.25ml	10% Triton X-100	0.25% Triton X-100
32.25ml	dH_2O	

store at room temperature

2.4.2.2 Lysis Buffer-2 (LB2)

		Final concentration
0.5ml	1M Tris-HCl pH8.0	10mM Tris
2.0ml	5M NaCl	200mM NaCl
100µl	0.5M EDTA	1mM EDTA
50µl	0.5M EGTA	0.5mM EGTA
47.35ml	dH_2O	

store at room temperature

2.4.2.3 Lysis Buffer-3 (LB3)

Final concentration

0.5ml	1M Tris-HCl pH8.0	10mM Tris
1.0ml	5M NaCl	100mM NaCl
100µl	0.5M EDTA	1mM EDTA
50µl	0.5M EGTA	0.5mM EGTA
0.5ml	100% Na-Deoxycholate	0.1% Na-Deoxycholate
1.25ml	20% N-lauroylsarcosine	0.5% N-lauroylsarcosine
46.60ml	dH_2O	

store at RT

2.4.3 Chromatin Immunoprecipitation (ChIP)

Following sonication, samples were centrifuged at 16,000g for 25 minutes at 4°C to pellet any cellular debris. The supernatant (~100µl), containing sonicated chromatin, was then transferred in to a clean DNA-low Bind tube, followed by addition of 10µl of 10% Triton X-100 (1% final concentration) to facilitate antibody stability by sequestering anionic detergent. 2.5µg of appropriate ChIP-seq grade antibody was then added and the samples incubated at 4°C for 12-16 hours (overnight), on a vertical spinning wheel, to bind antibody to chromatin.

Following overnight incubation, the antibody-bound chromatin was transferred to a clean DNA-low bind tube containing bovine serum albumin (BSA)-blocked 40µl Protein A Dynabeads (Invitrogen). This Dynabead-antibody-chromatin mixture was then incubated at 4°C for 3 hours on vertical spinning wheel to bind the Dynabeads to antibody. The Dyanbead-antibody-chromatin mixture was then washed 5 times with 1ml ice-cold RIPA wash buffer to remove any unbound antibody, chromatin, and other cellular debris and to reverse any non-specific binding between antibody and chromatin. Following the final wash, the beads are finally washed with 1ml ice-cold TE buffer containing 50mM NaCl. The resulting cleaned Dynabead-antibody-chromatin complex, containing antibody enriched chromatin, was re-suspended in

210µl of Elution Buffer and incubated at 65°C for 15 minutes, with continuous shaking at 850rpm, to release the antibody-chromatin complex from Dynabeads. The sample in Elution Buffer was then cleared off of Dynabeads and transferred to a clean DNA-low Bind tube and incubated at 65°C overnight to reverse formaldehyde mediated cross-linking of DNA with histone proteins.

The de-crosslinked DNA-protein mixture was then treated with Proteinase K (Invitrogen) to a final concentration of 0.2mg/ml and incubated at 55°C for 2 hours to digest histone proteins. Enriched DNA was subsequently purified using Zymo Research's ChIP DNA Clean and Concentrator kit as per the manufacturer's instructions. Samples were eluted in 6µl for use in qPCR or 10µl for sequencing library preparation.

2.4.3.1 Block Solution

0.5% BSA (w/v) in sterile PBS, stored at 4°C.

2.4.3.2 RIPA Wash Buffer

		Final Concentration
12.5ml	1M Hepes-KOH pH7.5	50mM Hepes
25.0ml	5M LiCl	500mM LiCl
0.5ml	0.5M EDTA	1mM EDTA
25.0ml	10% NP-40 (or substitute)	1% NP-40
17.5ml	10% Na-Deoxycholate	0.7% Na-Deoxycholate
169.5ml	dH_2O	

store at room temperature

2.4.3.3 Elution Buffer

		Final Concentration
2.5ml	1M Tris-HCl pH8.0	50mM Tris
1.0ml	0.5M EDTA	10mM EDTA
5.0ml	10% SDS	1% SDS
41.5ml	dH_2O	

store at room temperature

2.4.4 Real Time PCR using SYBR-Green dye

Enrichment for positive and negative control loci in ChIP samples was tested by qPCR using SYBR Green chemistry (Roche). 7.5µl of SYBR Green reaction mix was mixed with 2.5µl of diluted DNA in a well of a multiwell qPCR plate and PCR carried out as per manufacturer's instructions.

Each SYBR Green reaction mix contained:

5µl 2X SYBR Green Master Mix (Roche)

0.5µl 10µM Forward primer

0.5µl 10µM Reverse primer

1.5µl water

2.4.5 Sequencing Library Preparation

Sequencing libraries were prepared from ChIP-ed DNA using Diagenode's MicroPlex Library Preparation kit, as per the manufacturer's instructions. Step C.5 in the library preparation protocol was performed using the LightCycler with 1µl of PicoGreen dye (Life Technologies) to visualize amplification curves. Sequencing libraries were store at -20°C.

2.4.6 Sequencing Library Quantification

Sequencing libraries, before and after purification, were quantified using Quant-iT PicoGreen dsDNA Kit (Life Technologies) as per the manufacturer's instructions. Briefly, 1µl of library was diluted to 100µl using sterile TE buffer (pH8.0 was used instead of pH7.5 as recommended by the manufacturer) and mixed with equal volume of 1:200 diluted PicoGreen dye in a well of 48-well flat-bottom Corning plate. The resulting 200µl mixture was incubated at room temperature for 5 minutes and subsequently quantified using FLUOstar Omega (BMG Labtech). DNA standards were prepared in similar way using the Lambda DNA provided in Quant-iT PicoGreen dsDNA Kit.

2.4.7 Sequencing ChIP libraries on HiSeq 2000

The ChIP libraries were sent to Source Bioscience, a commercial next generation sequencing service, for sequencing on a HiSeq 2000. Libraries were multiplexed in one lane, generating a total of 120M 50-bp single end reads.

2.4.8 ChIP-seq data analysis

The sequencing data generated for ChIP libraries was analyzed using GeneProf (Halbritter et al. 2011), an analysis software produced by Tomlinson Lab, University of Edinburgh. The raw sequenced reads were aligned to NCBIM37/mm9 mouse reference genome using Bowtie algorithm (Langmead et al. 2009), using default GeneProf parameters: mean quality score ≥10, max mismatch = 2. Peaks were called from the aligned reads using MACS algorithm (Zhang et al. 2008; Feng et al. 2012), using default GeneProf parameters: bandwith = 100, PET inset size = 200, p-value ≤ 10^{-5}, fold enrichment over background (for model) = 10-30, local lamda = 1000-10000, FDR < 0.1. Identified peaks were then used for downstream analysis.

2.5 Tissue Culture

2.5.1 Thawing of Frozen Embryonic Stem (ES) Cells

ES cells were routinely stored in liquid nitrogen for long-term storage. To thaw ES cells, a T25 tissue culture flask (Corning) was coated with pre-warmed 0.1% gelatin for 15 minutes. Cells were retrieved from liquid nitrogen and quickly thawed in a 37°C water bath. Thawed cells were transferred to a universal tube containing 9.5ml of pre-warmed ES cells media, to dilute DMSO (see section 2.5.3), and centrifuged at 1000rpm for 5 minutes in a bench top centrifuge. The supernatant was carefully aspirated and the cell pellet gently re-suspended in 1ml of pre-warmed ES cell media. The excess gelatin in the T25 flask was removed and replaced with 6-7ml of pre-warmed ES cell media. Thawed ES cells were transferred to media-containing flask and placed in a 37°C incubator with 7.5% CO_2. Media was change either at the end of the day, if cells were thawed first thing in the morning, or in the morning of the next day.

2.5.2 Passaging Cells

To passage the cultured cells, the media was aspirated and the cells washed twice with sterile pre-warmed PBS. Cells were then treated with appropriate volume of trypsin (0.025% trypsin in PBS) or trypsin-EDTA (1.3mM EDTA) and placed in 37°C incubator for between 0.5-3 minutes (some primary cells require longer incubation) to detach cells. Pre-warmed media was then added, 4 times the volume of trypsin used, to neutralize trypsin and pipetted gently to dissociate colonies in to single cell suspension. The cell suspension was then transferred to a sterile universal tube and centrifuged at 1200rpm for 3 minutes in a bench top centrifuge. The supernatant was discarded and cell pellet re-suspended in 1ml of media. Cells were diluted as required and transferred to a fresh pre-coated flask or plate and returned to cell culture incubator.

2.5.3 Freezing ES Cells

To freeze ES cells, cells were harvested as for regular cell passaging. The cell pellet after dissociation was re-suspended in ES cell media containing 10% DMSO

(freezing solution) and then centrifuged at 1000rpm for 5 minutes. The supernatant was discarded and cell pellet re-suspended in appropriate volume of freezing solution. 0.5ml of this cell suspension was added to each 1ml cryotube vial. The number of cryotube vials produced depended on the size of the flask used for growing ES cells. Typically, 4-5 vials were produced from a near confluent T75 flask and 2-3 vials produced from a T25. Cryotubes were stored in -80°C freezer overnight after which they were transferred to liquid nitrogen cell bank.

2.5.4 Culturing Embryonic Thymic Tissue

Thymic tissue was routinely cultured in 48-well tissue culture plates using N2B27 media supplemented with 1:100 Pen/Strep antibiotics. Cytokines were added to media as required and media was changed everyday to prevent major changes in cytokine concentration throughout culture duration. Briefly, dissociated embryonic thymic tissue was obtained as described in previous sections. The tissue was washed twice with N2B27 containing antibiotics and transferred to tissue culture hood. The cell pellet was re-suspended in required volume of media and added to tissue culture plate pre-coated with Matrigel. Cells were transferred to 37°C incubator with 5% CO_2.

2.5.4.1 Coating Tissue Culture Plates with Matrigel

The tissue culture plates used to culture embryonic thymic tissue were coated with Matrigel (Corning). Matrigel was thawed and kept on ice through the procedure. Tissue culture plates and pipette tips were cooled to 4°C before use. Adding matrigel to each well and placing the plate at room temperature for 15-20 seconds to allow polymerization of matrigel created a thin layer of matrigel. The remaining matrigel was aspirated and the cell suspension added to the well.

2.5.5 ES cell media

1X Glasgow Minimum Essential Medium (Invitrogen)

10% foetal calf serum

1X non-essential amino acids

4mM glutamine

2mM sodium pyruvate

0.1mM 2-mercaptoethanol

1X leukaemia inhibitory factor (LIF)

2.5.6 N2B27 media

1X DMEM/F12-N2 with 1X Neurobasal/B27 medium in 1:1 ratio (SCS)

0.1mM β-mercaptoethanol (Sigma)

1μM PD (Stem Cell Sciences)

10ng/ml LIF

0.005μg/ml Recombinant Human BMP4 (R&D systems)

2.6 Molecular Techniques: Cloning

2.6.1 TOPO TA Cloning

Full-length cDNA was amplified for desired gene using either Phusion High-Fidelity DNA Polymerase (Thermo Scientific) or PrimeSTAR HS DNA Polymerase (Clontech) as per the manufacturers' instructions. The PCR product was checked by gel electrophoresis using 1.5% agarose gel. The band representing amplified PCR product was cut out of the gel and cleaned up using QIAquick Gel Extraction Kit (Qiagen). The purified full-length PCR product was then incubated at 72°C for 30 minutes with 10X PCR buffer, dATPs, and Taq polymerase to add dATPs to the 3` of the product, which allows the product to be ligated to TOPO vectors. The resulting PCR product was purified using QIAquick PCR Purification kit (Qiagen) and 50ng of the purified product was then mixed with 10ng of pCR 2.1-TOPO vector (1μl),

1µl salt solution, and water to a total volume of 6µl. The cloning reaction was incubated at room temperature for 30 minutes. 2µl of the cloning reaction was subsequently used for transformation of either TOP10 competent cells (Invitrogen) or DH10B competent cells (NEB) according to manufacturers' instructions for chemical transformation.

2.6.2 Restriction Enzyme Digestion

Restriction digests were carried out according to the manufacturer's guidelines. Typically, digests were carried out in a 50µl total reaction volume with 1U of enzyme per 1µg of DNA in 1X NEB buffer. Reactions were supplemented with 100µg/ml BSA if required. Reactions were typically incubated at 37°C for 2-3 hours, following which the enzyme was inactivated and the digestion products separated by gel electrophoresis.

2.6.3 Agarose Gel Electrophoresis

Gel electrophoresis was carried out as described above.

2.6.4 Gel Extraction

PCR products and restriction digests products were separated by agarose gel electrophorosis to allow isolation of single bands. The bands representing the desired products were excised using a sterile scalpel (Swann-Morton) under a UV transilluminator (Herolab, GmbH Laborgerate) that emits at a wavelength of 254nm. To reduce damage to DNA from excess UV exposure, a Perspex plate was placed between the transilluminator and the gel. Cut gel slices were transferred to a 1.5ml microcentrifuge tube. Extraction of DNA from gel slices was carried out using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's guidelines.

2.6.5 Ligation reaction

Isolated DNA fragments were ligated into vectors using T4 DNA Ligase enzyme (NEB) as per the manufacturer's guidelines. Briefly, digested vector and insert (DNA fragment) were mixed in 1:5 molar ratio in presence of 1µl of T4 DNA Ligase Buffer, 1µl T4 DNA Ligase enzyme, and water to a final volume of 20µl. A vector only control was set up to determine vector background. The ligation reaction was incubated at room temperature for 12-16 hours, following which 2µl was used to transform competent bacteria.

2.6.6 Plasmid DNA isolation

Transformed bacteria were plated on agar plates containing suitable antibiotics for selection of bacterial clones containing plasmid. Typically, 100µg/ml ampicillin was used. A single bacterial colony containing transformed plasmid was used to set up each starter culture by transferring to a 15ml round bottom polystyrene tubes containing 5ml Luria-Bertani (LB) culture medium with antibiotics and followed by incubation at 37°C for 8 hours with vigorous shaking (~250rpm) to allow bacterial growth. The starter cultures were then diluted 1/500 into 100ml fresh selective LB medium and grown overnight. Plasmid purification was carried out from these cultures using the Plasmid Maxi Kit (Qiagen) according to the manufacturer's protocol. Plasmid DNA was re-suspended in the Elution Buffer provided with the kit. DNA concentration and purity were measured using the NanoDrop ND-1000 Spectormeter (Labtech) and samples stored at -20°C until required.

2.6.7 Luria-Bertani (LB) Broth

10g Bacto-tryptone (BD Biosciences)

5g Bacto-yeast extract (BD Biosciences)

10g Sodium Chloride (NaCl)

Make up to 1L with deionized water and adjust pH to 7.5 with sodium hydroxide (NaOH) and autoclave.

2.6.8 Transfection of ES cells with Lipofectamin

Lipofectamin 2000 (Invitrogen) was used to transfect ES cells with plasmids. All steps were carried out in a sterile environment inside a tissue culture hood. Transfection was typically carried out in 6-well plates. ES cells were harvested as described above and 10^6 cells were plated in one well of a pre-coated 6-well plate containing ES cells media. Transfection reagents were prepared according to the manufacturer's instructions. 3µg of plasmid DNA was mixed with 250µl of ES cells media without FCS. Similarly, 3µl of Lipofectamine 2000 was mixed with 250µl of serum free ES cells media in a separate microcentrifuge tube. Both plasmid and lipofectamine were incubated at room temperature for 5 minutes, following which media containing plasmid was mixed with that containing lipofectamine, mixed by gentle pipetting, and incubated at room temperature for 20 minutes. This enables the plasmid DNA to be encapsulated by lipofectamine reagent. This lipofectamine-plasmid mixture was then added to the well containing freshly plated ES cells. The transfected ES cells were cultured as described above. The media was replaced with fresh ES cells media on the morning after the transfection. ES cells were cultured for 48 hours after transfection before use.

2.7 Molecular Techniques: RNA-seq

The protocol for RNA-seq was obtained from Prof. Sten-Erik Jacobsen's lab at the University of Oxford. RNA-seq was performed using SMART-seq2 protocol (Picelli et al. 2014). Only sample collection was performed in our lab. The samples were then sent to Jacobsen lab for processing.

2.7.1 RNA-seq sample collection

Before sample collection for RNA-seq, all the equipments - pipettes, pipette tips, DNase and RNase-free 1.5ml microcentrifuge and 0.2ml PCR tubes, marker pens, and clean hood - were treated with disinfectant, DNA-OFF (Takara Bio), and

RNaseZap (Ambion) and treated with UV light for at least 30 minutes. Reagents were treated with disinfectant, DNA-OFF, and RNaseZap every time before placing in to the clean hood. Reagent preparation was carried out away from the regular lab space to avoid contamination. The personal protective equipment (PPE) used for RNA-seq reagent preparation and sample collection were kept separate from the PPE routinely used in the lab. PPEs included disposable lab coat, apron, sleeves, and gloves, which were discarded after every use.

The flow cytometric cell sorter, Aria-II (BD Bioscience), was cleaned and sterilized before use. Cells were sorted as per the normal protocol. 100 cells were sorted into clean, sterile 0.2ml PCR tubes containing 5µl of reaction mix. Sample were centrifuged briefly and placed on dry ice immediately. Samples were then transferred to -80°C until required.

2.7.1.1 Cell Lysis Buffer for RNA-seq

19µl 0.2% Triton X-100 (Sigma)

1µl Recombinant RNase inhibitor (Clontech)

2.7.1.2 Reaction Mix for Sample Collection

Each 0.2ml PCR tube contains:

2µl Cell Lysis Buffer

1µl 10µM oligo-dT primer (custom made, see (Picelli et al. 2014))

1µl dNTP mix (10mM each)

2.8 CRISPR-Cas9 for genome editing

The CRISPR-Cas9 system was used to delete genomic regions containing binding sites for transcription factor of interest.

2.8.1 guide-RNA (gRNA) Design

gRNAs were designed using E-CRISP (http://www.e-crisp.org/E-CRISP/) online software (Heigwer et al. 2014). The sequence for genomic region containing the binding site(s) for transcription factor(s) was obtained from ENSEMBL and used for designing gRNAs. Default parameters were used to design gRNAs, except for parameters in "Off-Target analysis".

Additional base pairs were added to the 5` and 3` of gRNAs and their complementary sequences to allow for direct cloning in to BbsI digested plasmid.

2.8.2 Annealing gRNA and its Complementary Sequence

gRNAs and their complementary sequences were obtained from Sigma. The complementary sequences were annealed together to obtain double-stranded DNA, which could be directly ligated to BbsI digested expression vector. Annealing reaction mix was heated to 98°C to denature oligos and then gradually cooled to 20°C over 2 hours and 39 minutes to allow for annealing of complementary oligos. Annealed oligos were diluted 1:500 (~100nM) using TE buffer and 4µl of diluted oligos was ligated to the expression vector.

Annealing reaction mix:

9µl 100µM sense oligo

9µl 100µM antisense oligo

2µl 10X T4 DNA ligase buffer

2.8.3 Ligation of Annealed Oligos and Vector

Annealed oligos were ligated to BbsI digested vector to obtain vector containing Cas9 and gRNA sequences. Ligation reaction mix was incubated at room

temperature for 2 hours. Ligated vector was used for transforming DH10B competent cells as described above.

Ligation reaction mix:

4µl 100nM annealed oligos

1µl BbsI cut vector

1µl 10X T4 DNA ligase buffer

1µl T4 DNA ligase enzyme

3µl dH_2O

2.9 Statistics

2.9.1 Statistical test for qPCR data

Statistical significance for qPCR data was calculated using Student's T-test. First, the expression value of a gene in each sample was converted to logarithmic scale to enable statistical testing for genes with low expression levels. Biological replicates were treated as separate sample within a group. Student's T-test was run using one-tailed distribution for samples with unequal variance.

2.10 Tables

2.10.1 Genotyping primers

Primer	**Sequence**
β-gal F	TAATGGGATAGGTTACGT
β-gal R	ACAACGTCGTGACTGGGAAAA
Foxn1 exon-3 F	CTCCAGAGAGGACACCCTCA

Foxn1 exon-3 R	CTCTGCTGGGAAGCTAGGC

2.10.2 Primary antibodies for FACS

Antibody	Clone	Isotype	Conjugate	Company	Concentration
EpCAM	G8.8	Rat IgG2a	PE	Biolegend	0.25µg/ml
CD45	30-F11	Rat IgG2b	APC, PerCp/Cy5.5	BD/eBioscience	1µg/ml
CD31	390	Rat IgG2a	APC, PerCp/Cy5.5	BD/eBioscience	1µg/ml
Ter119	Ter119	Rat IgG2b	APC, PerCp/Cy5.5	BD/eBioscience	1µg/ml
Ly51	6C3	Rat IgG2a	Biotin	Biolegend	5µg/ml
MHC II	M5/114.1 5.2	Rat IgG2b	PE	BD	1µg/ml
CD11b	M1/70	Rat IgG2b	FITC	BD	2.5µg/ml
CD11c	HL3	Hamster IgG1	FITC, PerCP/Cy5.5	BD	2.5µg/ml
Plet1	MTS24	Rat IgG2a	-	R.L.Boyd, Monash University, Melbourne	Neat
UEA1		Rat IgG2a	-		

2.10.3 Secondary antibodies for FACS

Flurochrome	Isotype	Company	Concentration
Alexa568	IgGa/b	Invitrogen	4µg/ml
Alexa647	IgGa/b	Life Technologies	4µg/ml

Alexa488	IgG2a/b	Molecular Probes	4µg/ml

2.10.4 ChIP-seq antibodies

Antibody	**Source-Isotype**	**Company**
H3K4me3	Rabbit IgG	Diagenode (pAb-003-050)
H3K27ac	Rabbit IgG	Abcam (ab4729)
H3K4me	Rabbit IgG	Abcam (ab8895)
panH3	Rabbit IgG	Abcam (ab1791)
IgG	Rabbit IgG	Diagenode (kch-504-250)

2.10.5 Cytokines for TEPC culture

Mouse TGFβ1 Recombinant Protein (eBioscience; Catalogue number: 14-8342)

2.10.6 ChIP-qPCR primers

Primer	**Sequence**
Esrrb 3` F	ACTCCTCCCCTTACCCCTGT
Esrrb 3` R	GGCTGTGGTCACTGCATCTA
Esrrb E6 F	CCCATTTTGTACCAGTCCTTGA
Esrrb E6 R	TAGCCAGGTAGCCATCCAAA
Nanog enh-2 F	GATCTTGTCTCCTGCGTGCT
Nanog enh-2 R	TTGGCACAGTACTCGCTTTG
Gapdh pro F	Diagenode (pp-1045-050)
Gapdh pro R	Diagenode (pp-1045-500)
FN1.1 F	ACCATGCTTCAGCCAGACTC
FN1.1 R	GGACAGGTGGGTTGTATGGG
FN1.4 F	AGTGTGGGATCTGTGTGTGG
FN1.4 R	CCAGTGAGTGAGCCATGGAA

FN1.7 F	ATTGCCTCCCAACTCAAGGT
FN1.7 R	GCCCAGGCCTTCTGACTTT
FN1.10 F	CCTCGCAGGGATTTGTCTGT
FN1.10 R	AATGACACCGTAGCGTGGAG
FN1.13 F	ACATCCTATGGCGCTGAACC
FN1.13 R	GCCTGCAGACAGATGGAAGT

2.10.7 Quantitative real-time PCR primers

Name	**Sequence**	**UPL Probe #**
Pax1 F	CTCCGCACATTCAGTCAGC	105
Pax1 R	TCTTCCATCTTGGGGGAGTA	
Pax9 F	AGCAGGAAGCCAAGTACGG	33
Pax9 R	TGGATGCTGAGACGAAACTG	
Foxn1 F	TGACGGAGCACTTCCCTTAC	68
Foxn1 R	GACAGGTTATGGCGAACAGAA	
Hoxa3 F	CAAGGCAGAACACTAAGCAGAA	78
Hoxa3 R	GTCACCAGCGCAGCTCTC	
Foxg1 F	GAAGGCCTCCACAGAACG	26
Foxg1 R	CAAGGCATGTAGCAAAAGAGC	
p63 F	TGGAAAACAATGCCCAGACT	45
p63 R	CTGCTGGTCCATGCTGTTC	
Dll4 F	AGGTGCCACTTCGGTTACAC	106
Dll4 R	GGGAGAGCAAATGGCTGATA	
CCL25 F	GAGTGCCACCCTAGGTCATC	9
CCL25 R	CCAGCTGGTGCTTACTCTGA	
Smad7 F	ACCCCCATCACCTTAGTCG	63
Smad7 R	GAAAATCCATTGGGTATCTGGA	

Pth F	CAGTCCAGTTCATCAGCTGTC	78
Pth R	GTGTTTGCAGACATCATCTTTACAT	
Tgfb2 F	TCTTCCGCTTGCAAAACC	106
Tgfb2 R	GTGGGAGATGTTAAGTCTTTGGA	
HMBS F	TCCCTGAAGGATGTGCCTAC	79
HMBS R	AAGGGTTTTCCCGTTTGC	
b-actin F	AAGGCCAACCGTGAAAAGAT	56
b-actin R	GTGGTACGACCAGAGGCATAC	
HPRT F	TCCTCCTCAGACCGCTTTT	95
HPRT R	CCTGGTTCATCATCGCTAATC	
Hey1 F	AAGAGAGCTCACCCAGACTAC	17
Hey1 R	GAACACAGCGCCGAACTC	
Thap11 F	CTGGTGGTGGTAGGGGAAG	21
Thap11 R	CCCGAGGACAAGGAGTACG	
Eif3a F	GTCCTTGGAGACCATTTGTCA	58
Eif3a R	GTGGACCCCAACTCTCTTCTC	
Bhlhe40 F	GAAGGATCTCCTACCCGAACA	50
Bhlhe40 R	GCTTTCACGTGCTTCAACG	
Tax1bp3 F	GAGGGAGGTCCTGCTGAAAT	26
Tax1bp3 R	GTCCCAGCCATTCACCTG	
Ing4 F	ATTGCCTTTGTCACCAGGTC	76
Ing4 R	GGAACCACTCGATGGAACAA	
Six4 F	CCAGTATGGCATTGTCCAGAT	50
Six4 R	CGAAGTGCTTGGGGTAACTG	
Zfp503 F	GAAGCACCCAAAGGTAGAAGG	62
Zfp503 R	AACGCAGGCGGTAATGTG	
Yap1 F	CAGGAATTATTTCGGCAGGA	71
Yap1 R	CATCCTGCTCCAGTGTAGGC	

Irf6 F	AAGAGTGGCCAGATGGAAAG	85
Irf6 R	ATCATGCGAGCCACCACT	
Creb3l2 F	GCTGGAGGACACCAATAGGA	9
Creb3l2 R	AGCTAACTTGCAGGTTCGAGAG	
Foxc1 F	ACCACGTAAGTTTCTTGCGTTC	15
Foxc1 R	GAGGCAGAGGGGCAAGAC	
Hes6 F	GGATCAACGAGAGTCTTCAGGA	66
Hes6 R	TTCTCTAGCTTGGCCTGCAC	
Eid1 F	TGATCGAGAGTAAATGCTGACG	101
Eid1 R	CCCAAAATAGCAGAATGTTGAA	
Zfp36l1 F	CCATTTTTGACTTGAGCGAAG	27
Zfp36l1 R	GGCAGAGTGACCGAGTGC	
Foxo1 F	TGGGTGTCAGGCTAAGAGTTAGTC	68
Foxo1 R	AGGGGTGAAGGGCATCTTT	
NfiB F	CCCAAGATTGAGCACTTTCC	109
NfiB R	GGAGGTGGAGTTCGAGTTGA	
Six2 F	CAAGTCAGCAACTGGTTCAAGA	5
Six2 R	ACTGCCATTGAGCGAGGA	
Foxa1 F	GAACAGCTACTACGCGGACA	82
Foxa1 R	CGGAGTTCATGTTGCTGACA	
Foxa2 F	AAGTAGCCACCACACTTCAGG	32
Foxa2 R	TGGCCCATCTATTTAGGGAC	
Tcf3 F	CGTCATCCTCAGCCTGGA	52
Tcf3 R	ACCACGCCAGACACCTTC	
p53 F	ATGCCCATGCTACAGAGGAG	78
p53 R	AGACTGGCCCTTCTTGGTCT	
Fbxw7 F	CTCAGACTTGTCGATACTGGAGAA	73
Fbxw7 R	GATGTGCAACGGTTCATCAAT	

Gata3 F	TTATCAAGCCCAAGCGAAG	108
Gata3 R	GTGGTGGTCTGACAGTTCG	
Kitl F	TCAACATTAGGTCCCGAGAAAG	71
Kitl R	ACTGCTACTGCTGTCATTCCTAAG	
SetDB1 F	CCAAAGGCTCTTTTGTCTGC	80
SetDB1 R	CAGATTTGCAAAGTACTCATCACC	
E2F3 F	CAAGGACCCTCCAGCAGAG	70
E2F3 R	AGTTCCAGCCTTCGCTTTG	
Fgfr2IIIb F	CCCTGCGGAGACAGGTAAC	17
Fgfr2IIIb R	CGGGGTGTTGGAGTTCAT	
Sox9 F	GTACCCGCATCTGCACAAC	66
Sox9 R	CTCCTCCACGAAGGGTCTCT	
Six1 F	ACCGGAGGCAAAGAGACC	63
Six1 R	GGAGAGAGTTGATTCTGCTTGTT	
Eya2 F	TTCAACCTTGCTGACACACAC	85
Eya2 R	CACGTGGATTTGGTCACAGT	
Tbx1 F	GCTGTGGGACGAGTTCAATC	104
Tbx1 R	ACGTGGGGAACATTCGTCT	
Fgf8 F	CAGTGTCCTGCCTAAAGTCACA	70
Fgf8 R	CGGCTGTAGAGCTGGTAGGT	
Gcm2 F	ACTTGATGGCAATGCAATTTT	29
Gcm2 R	GAAGCCATCTGTCTCTTGAGG	
Notch1 F	GGATGCTGACTGCATGGAT	93
Notch1 R	AATCATGAGGGGTGTGAAGC	
E2F1 F	TGCCAAGAAGTCCAAGAATCA	5
E2F1 R	CTTCAAGCCGCTTACCAATC	
Hoxa3 F	CAAGGCAGAACACTAAGCAGAA	78
Hoxa3 R	GTCACCAGCGCAGCTCTC	

Cxcl12 F	GGTTCTTCGAGAGCCACATC	21
Cxcl12 R	TTCTTCAGCCGTGCAACA	
Eya1 F	ACAAAAACAACGTGGGAGGT	78
Eya1 R	GCCAGGAGTCTGTGAGTGC	
Hey1 F	ACCATCGAGGTGGAAAAGG	72
Hey1 R	CTTCTCGATGATGCCTCTCC	
Smad2 F	AGGACGGTTAGATGAGCTTGAG	9
Smad2 R	GTCCCCAAATTTCAGAGCAA	
Tgfb3 F	CCCTGGACACCAATTACTGC	25
Tgfb3 R	TCAATATAAAGGGGGCGTACA	
Smad4 F	GCCCCGGCAGAGTCTAAC	60
Smad4 R	GAATACTGGCCGGCTGAC	
Tgfbr2 F	AGAAGCCGCATGAAGTCTG	69
Tgfbr2 R	GGCAAACCGTCTCCAGAGTA	
Tgfbr1 F	GCAGCTCCTCATCGTGTTG	70
Tgfbr1 R	AGAGGTGGCAGAAACACTGTAAT	
Tgfb1 F	TGGAGCAACATGTGGAACTC	72
Tgfb1 R	GTCAGCAGCCGGTTACCA	
Smad3 F	TCAAGAAGACGGGGCAGTT	78
Smad3 R	CCGACCATCCAGTGACCT	

3. Investigating the expression of candidate transcriptional regulators of *Foxn1* identified during thymus development

3.1 Introduction

Given the importance of *Foxn1* in thymus development, homeostasis, and regeneration, it is crucial to understand the transcriptional regulation of this gene and functions of the encoded protein. Studies from our lab and others have suggested that FOXN1 is a direct transcriptional regulator of *Dll4* and *CCL25* (Bleul & Boehm 2000; Tsukamoto et al. 2005; Nowell et al. 2011). A recent elegant study showed that the expression of *Scf (Kitl)* and *Cxcl12* in *Foxn1*$^{-/-}$ thymus is absent during development, suggesting that these genes are also likely to be regulated by FOXN1 (Calderón & Boehm 2012). This study further demonstrated that expressing these four genes together under the control of a *Foxn1* promoter element confers some thymus functionality on *Foxn1*$^{-/-}$ mice, although full functionality was not achieved (Calderón & Boehm 2012). This suggests that these four genes together are capable of compensating, to an extent, for loss of *Foxn1* by carrying out some of its downstream functions that are important for thymocyte differentiation. Thus, at least a subset of functions of FOXN1 is mediated by its direct or indirect regulation of these four genes in TECs. On the other hand, very little is known about the regulation of *Foxn1* expression in the thymus. Previous studies have indicated that a *Tbx1-Six1/Eya1-Fgf8* genetic pathway, similar to that found in Drosophila, is important for formation of 3PP (Guo et al. 2011). However, the genetic pathway including these genes in mice has been shown to operate differently than that in Drosophila (Zou et al. 2006). Similarly *Hoxa3*, *Pax1*, and *Pax9* have also shown to be important for 3PP formation (Manley & Capecchi 1998; Su & Manley 2000; Hetzer-Egger et al. 2002; Chojnowski et al. 2014). To date, distinguishing the roles of these genes in 3PP formation and subsequently for thymus fate-commitment and TEC biology has proved to be difficult as defects in 3PP formation are invariably linked to defects in the organs derived from it, i.e. thymus and parathyroid (Balciunaite et al. 2002; Wei & Condie 2011; Chojnowski et al. 2014). This has made it difficult to determine whether any of these genes play a role in regulation of *Foxn1* expression in thymus. It is now clear that *Foxn1* expression is initiated in the

absence of HOXA3 (Chojnowski et al. 2014), while our unpublished data indicate that in the absence of both *Pax1* and *Pax9*, *Foxn1* expression is not initiated in the ectopic and very hypoplastic thymus primordium that forms in these mutants. Thus, while the above genes may play a role in the transcriptional regulation of *Foxn1*, with the exception of *Pax1* and *Pax9*, no clear evidence yet links this network directly to *Foxn1*. Recently, it was shown that induced expression of *Tbx1* in the developing thymic rudiment leads to downregulation of *Foxn1* expression (Reeh et al. 2014). Furthermore, a study focusing on the role of E2Fs and Rb proteins in the thymus has showed that E2Fs can regulate the expression of *Foxn1* in adult thymus, possibly through directly binding to its promoter (Garfin et al. 2013). However, it is not yet clear if E2Fs also regulate *Foxn1* expression during thymus development. Finally, HOXC13 has been suggested to be involved in regulation of *Foxn1* expression in hair follicle, however I can not detect the expression of *Hoxc13* in cells of 3PP or subsequent TEPCs in E12.5 thymi (unpublished), suggesting that *Hoxc13* is not required for initiation of *Foxn1* expression in TEPCs.

To take an unbiased, independent approach, we therefore decided to identify candidate transcriptional regulators of *Foxn1* using a bioinformatics approach. To this end, in work performed prior to the start of my PhD, I generated a list of candidate transcriptional regulators by performing correlation coefficient analysis on a microarray based gene expression dataset generated in the Blackburn lab. In brief, genes with expression patterns showing significant correlation or anti-correlation with that of *Foxn1* were identified using Pearson Correlation Coefficient and bootstrap techniques. The rationale was that the expression of *Foxn1* should closely follow the expression of its positive transcriptional regulators; whereas expression of a negative transcriptional regulator should be inversely correlated with *Foxn1*. This approach identified a list of genes that were positively or negatively correlated with *Foxn1* expression across a limited number of TEC datasets. Details of this approach and list of candidate transcriptional regulators can be found in my Master's thesis. As a central aim of my thesis was to identify transcriptional regulators of *Foxn1*, as a starting point of my studies I therefore set out to seek evidence that one or more of

these genes might be involved in regulating *Foxn1* expression during thymus development.

Thus, in this chapter I describe the changes in expression of these genes during early mouse thymus organogenesis in WT and *Foxn1*$^{-/-}$ mutants, in particular detailing their expression over the time-point at which *Foxn1* expression in TEC is initiated. For completeness, genes known to be important for thymus biology and expected to be either upstream or downstream of *Foxn1* regulatory network, regardless of them coding for transcription factors or not, were also included in this analysis, as their expression has so far not been studied extensively.

3.2 Expression analysis of candidate *Foxn1* regulatory genes

3.2.1 Sorting Strategy for Collection of 3PP cells and TEPCs

Since the microarray datasets used for identifying candidate transcriptional regulators of *Foxn1* were generated from E12.5 and E15.5 thymi, we decided to study the expression of the candidate genes during earlier developmental time points, including time-points immediately before and immediately after initiation of *Foxn1* expression in the thymus. To do this, timed mating of mice was set-up as described in Materials and Methods, and embryos were collected every 12 hours starting at embryonic day 9.5 (E9.5, the time of formation of 3PP). The developmental stage of the embryos was determined by counting the number of somites, as described in Theiler's Mouse Development Atlas (Theiler 1989). The, 3PP was then isolated from E9.5, E10.5, E11.0, and the thymic rudiment from E11.5, E12.0, and E12.5 embryos. Isolated tissue was dissociated and 200 EpCAM^{+}PLET1^{+} cells from each sample were sorted directly into reaction mix (see section 2.3.3) as described in Figure 3.1. Briefly, the gating strategy first excluded out the dead cells using 4` 6-diamidino-2-phenylindole (DAPI). Viable cells were then gated to exclude cells expressing lineage (lin) specific markers for T-cells (CD45), endothelial cells (CD31), and mesenchymal cells (Ter119). These Lin^{-}DAPI^{-} cells were then analysed for the

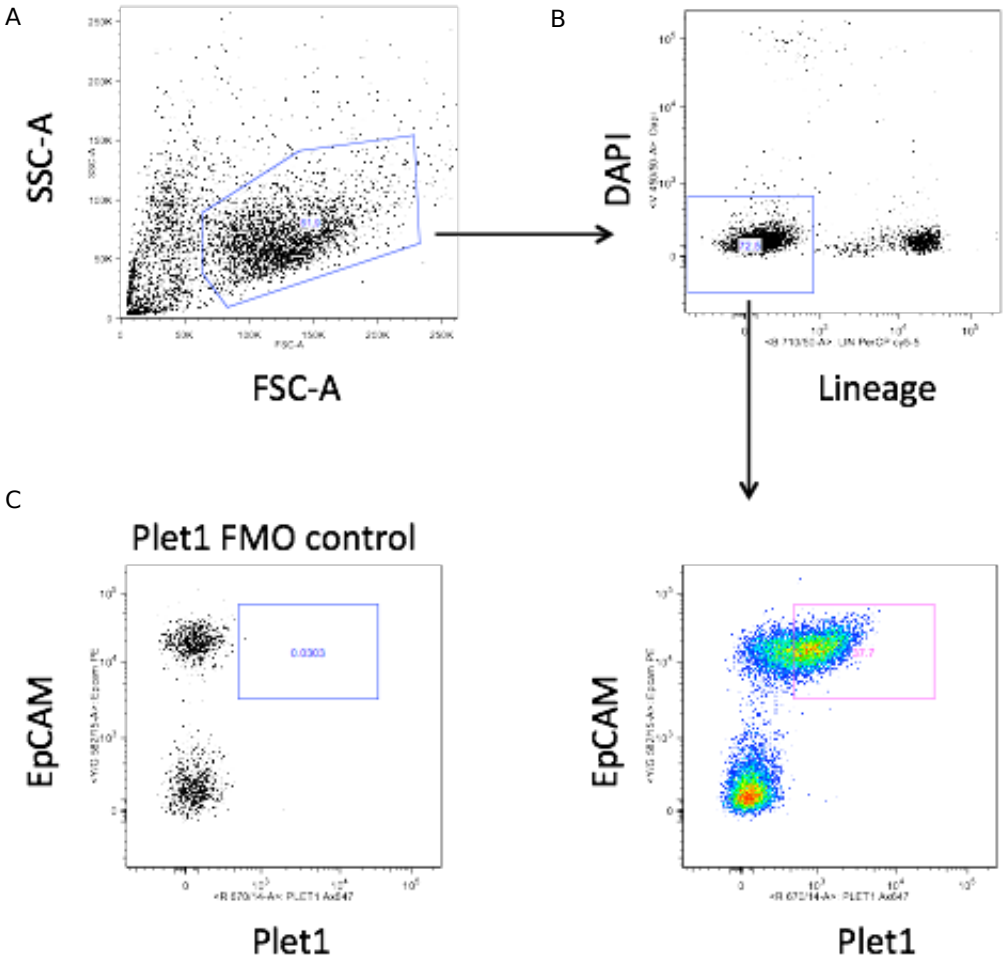


Figure 3.1: Sorting strategy for isolation of viable EpCAM^{+}Plet1^{+} third pharyngeal pouch endoderm cells and thymic epithelial progenitor cells.
(A) Gate showing 3PP/TEP cells gated based on size and density as determined based on forward scatter (FSC-A) and side scatter (SSC-A) of events regarded as single cells. (B) Gating of cells negative for dead cell marker DAPI and Lin (CD45: hematopoetic cells, CD31: endothelial cells, Ter119: red blood cells). (C) Gate for the isolation of EpCAM^{+}Plet1^{+} cells. Also shown is the FMO (full minus one - lacking primary antibody for Plet1) control for comparison.

presence of EpCAM and the TEPC marker PLET1 and cells positive for both these marker were isolated by flow cytometry. The sorted samples were then used for qPCR based gene expression analysis using the Fluidigm platform. An important caveat to this strategy is the presence of *Plet1* expressing parathyroid cells, as parathyroid and thymus are not separated during early developmental stages, in the sorted samples, which can influence the observed results.

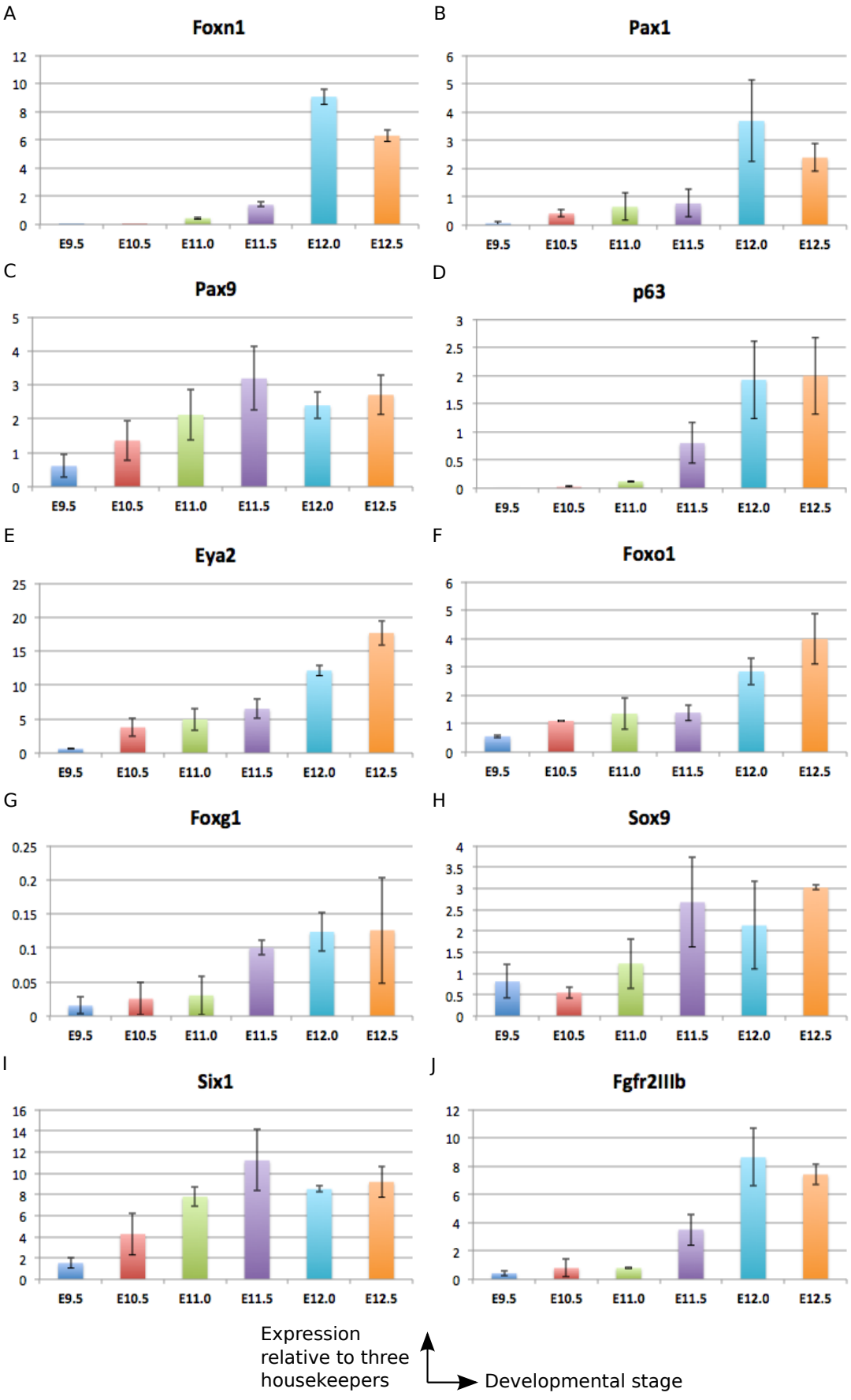
3.2.2 Gene Expression Changes in Developing 3PP and TEPCs.

The gene expression profiles generated from E9.5, E10.5, E11.0, E11.5, E12.0, and E12.5 showed two major patterns: 1) an increase in expression during development and 2) a decrease in expression during development or at E12.5. *Foxn1* expression was extremely low during the formation and early development of 3PP (expression relative to housekeeper: 0.017 (E9.5) and 0.077 (E10.5)), with higher level detected from E11.0 and then increased gradually till E12.5 (Figure 3.2A). The expression of most of the candidate transcriptional regulators of *Foxn1* showed a gradual increase in their expression across the developmental stages analyzed. These included genes such as *Pax1* (Figure 3.2B), *Pax9* (Figure 3.2C), *p63* (Figure 3.2D), *Eya2* (Figure 3.2E) *Foxo1* (Figure 3.2F), *Foxg1* (Figure 3.2G), *Sox9* (Figure 3.2H), and *Six1* (Figure 3.2I). Note that the expression of these genes is detectable prior to E11.5, the developmental stage at which *Foxn1* expression increases sharply throughout the thymic primordium.

Figure 3.2: Analysis of *Foxn1*, *Pax1*, *Pax9*, *p63*, *Eya2*, *Foxo1*, *Foxg1*, *Sox9, Six1*, and *Fgfr2IIIb* expression patterns during normal thymus organogenesis.

See following page

Relative expression levels in 3PP and TEP cells were determined by QRT-PCR. Graphs show the expression patterns in WT thymus for (A) *Foxn1*, (B) *Pax1*, (C) *Pax9*, (D) *p63*, (E) *Eya2*, (F) *Foxo1*, (G) *Foxg1*, (H) *Sox9*, (I) *Six1*, and (J) *Fgfr2IIIb*. Data are shown relative to the geometric mean Ct value for three housekeeping genes (*HMBS*, *b-actin*, *TBP*). Data shown are representative of two biological replicates and three technical replicates. Error bars show standard deviation (SD).


A
Foxn1
B
Pax1
C
Pax9
D
p63
E
Eya2
F
Foxo1
G
Foxg1
H
Sox9
I
Six1
J
Fgfr2IIIb
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
Expression relative to three housekeepers
Developmental stage

This is in accordance with the argument that a positive transcriptional regulator of *Foxn1* would be expressed prior to detectable levels of *Foxn1* expression. Furthermore, *Pax1*, *p63*, *Foxo1*, and *Eya2* showed increasing expressing from E11.5 to E12.5, similar to that observed for *Foxn1*, as would be expected of a positive transcriptional regulator of *Foxn1*. Thus, these genes represent strong candidates as transcriptional regulator of *Foxn1*. The expression of *Pax9*, *Foxg1*, *Sox9*, and *Six1* increased between E9.5 and E11.5, after which their expression remained relatively unchanged. A similar increase in expression was also observed for *Fgfr2IIIb* (Figure 3.2J), which is required for thymus development (Revest et al. 2001). *Fgfr2IIIb* was included in this analysis, as it has been suggested to be a target of FOXN1 (Bredenkamp, Nowell, et al. 2014), however its expression during development was unknown. The expression pattern of Fgfr2IIIb suggested that its expression is likely to be regulated by FOXN1 in the developing thymic rudiment. The expression of *Pax9*, *Foxg1*, *Sox9*, and *Six1* showed increased between E9.5 and E11.5, however their expression remained relatively constant thereafter. Whether the increased expression levels of these genes are important for initiation of high level expression of *Foxn1* can not be determined from the present data alone.

3.2.2.1 Foxa1 and *Foxa2*: genes important for endoderm development

The expression of *Foxa1* increased gradually from E9.5 to E11.0 after which its expression decreased until E12.5 (Figure 3.3A). The increase in expression of *Foxa1* until E11.0 suggests that *Foxa1* may play an important role in generation of 3PP endoderm and acquisition of competence to respond to subsequent commitment and differentiation signals, similar to its role in liver development (Lee et al. 2005). Beyond E11.0, *Foxa1* may be less important, or indeed detrimental, for further development of 3PP into thymus, as suggested by its declining expression. Interestingly, this is also the developmental time point when the expression of several genes important for thymic epithelial differentiation and function, such as *Foxn1* and its targets, is initiated. Thus, it is likely that the endodermal gene programme in 3PP endoderm is replaced by thymic epithelial cells specific gene

programme around E11.0. The expression level of *Foxa2* in PLET1^{+} TEC (Figure 3.3B) was much lower than that of *Foxa1* at all the time points analysed, suggesting that *Foxa1* is potentially more important in 3PP endoderm. While the expression level of *Foxa2* varied at different time points, the presence of considerable biological variations (as represented by the error bars) made it difficult to interpret these results.

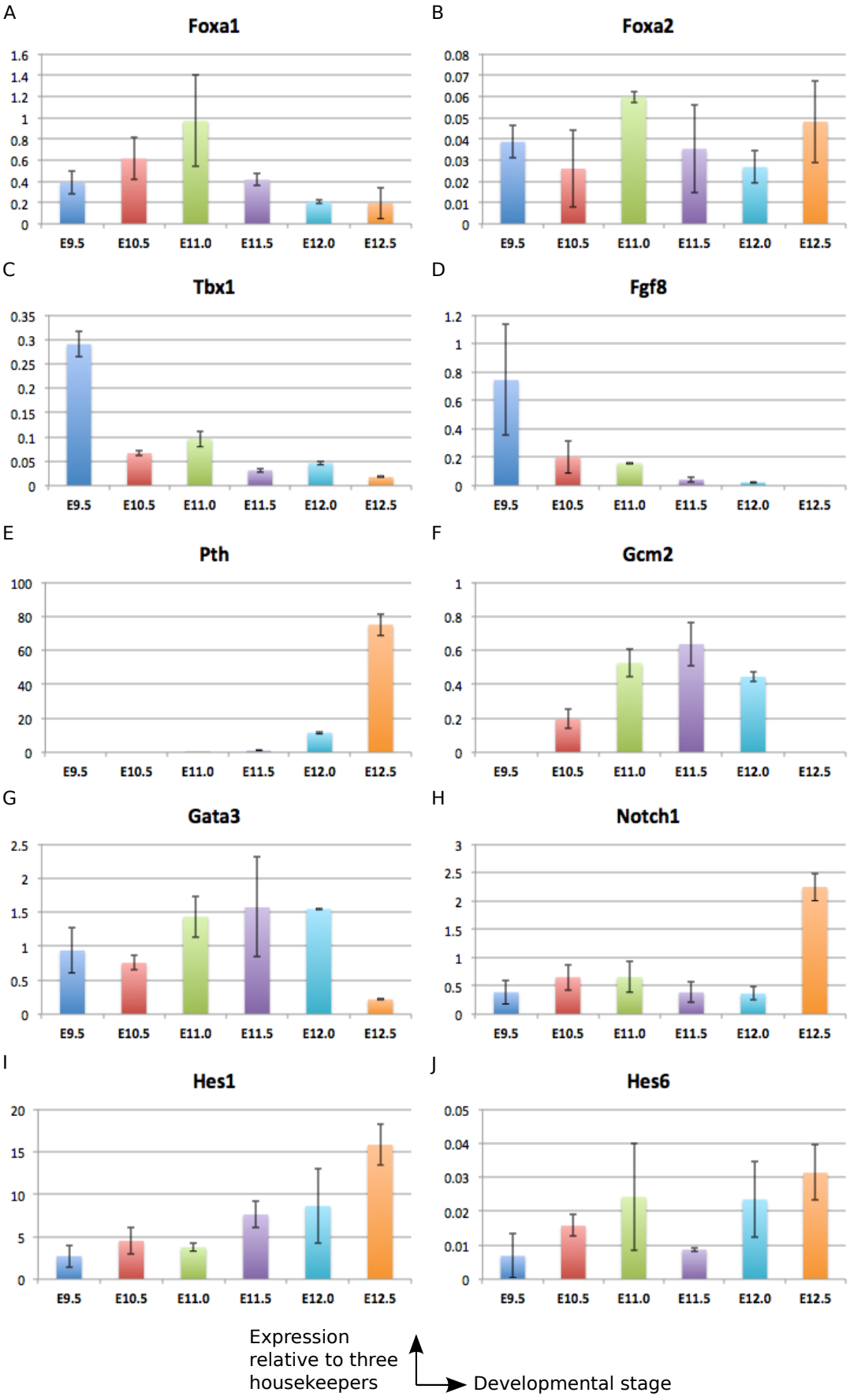
3.2.2.2 Tbx1 and *Fgf8*: genes important for pharyngeal arch and pouch formation

Tbx1 has been shown to be important for the formation of pharyngeal arches and pouches (Jerome & Papaioannou 2001; Xu et al. 2005; Reeh et al. 2014). Consistent with this, *Tbx1* expression was substantially higher at E9.5 than at other stages of thymus development and showed a 16-fold reduction between E9.5 and E12.5 (Figure 3.3C). The temporal regulation of *Tbx1* expression has previously been shown to be important for normal pharyngeal development (Okubo et al. 2011). The decrease in *Tbx1* expression after E9.5 is consistent with published data showing that induced expression of this gene after E11.5 is detrimental for thymus development (Reeh et al. 2014). Such induced expression of *Tbx1* in thymic epithelial cells has been shown to result in decreased *Foxn1* expression in these cells (Reeh et al. 2014).

Figure 3.3: Analysis of *Foxa1*, *Foxa2*, *Tbx1*, *Fgf8*, *Pth*, *Gcm2*, *Gata3*, *Notch1*, *Hes1*, *and Hes6* expression patterns during normal thymus organogenesis.

See following page

Relative expression levels in 3PP and TEP cells were determined by QRT-PCR. Graphs show the expression patterns in WT thymus for (A) *Foxa1*, (B) *Foxa2*, (C) *Tbx1*, (D) *Fgf8*, (E) *Pth*, (F) *Gcm2*, (G) *Gata3*, (H) *Notch1*, (I) *Hes1*, and (J) *Hes6*. Data are shown relative to the geometric mean Ct value for three housekeeping genes (*HMBS*, *b-actin*, *TBP*). Data shown are representative of two biological replicates and three technical replicates. Error bars show standard deviation (SD).


A
Foxa1
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
B
Foxa2
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
C
Tbx1
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
D
Fgf8
1.2
1
0.8
0.6
0.4
0.2
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
E
Pth
100
80
60
40
20
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
F
Gcm2
1
0.8
0.6
0.4
0.2
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
G
Gata3
2.5
2
1.5
1
0.5
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
H
Notch1
3
2.5
2
1.5
1
0.5
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
I
Hes1
20
15
10
5
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
J
Hes6
0.05
0.04
0.03
0.02
0.01
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
Expression relative to three housekeepers
Developmental stage

Whether the low level *Tbx1* expression observed till E11.0 is sufficient to inhibit initiation of *Foxn1* expression remains to be determined. Studies on DiGeorge syndrome have previously shown the importance of *Tbx1-Six1/Eya1-Fgf8* pathway in cardiovascular and craniofacial morphogenesis (Guo et al. 2011) and a possible *Tbx1-Fgf8* pathway also exists in 3PP development as demonstrated by a defect in 3PP development in *Fgf8* hypomorphic mutants (Abu-Issa et al. 2002; Frank et al. 2002). Therefore, I decided to also analyse the expression of *Fgf8* during thymus development. Similar to *Tbx1*, the expression of *Fgf8* is highest at E9.5 and decreases significantly (p-value = 0.016 (vs E11.5); p-value = 0.023(vs E12.0)) during development (Figure 3.3D), such that its expression was not detected at E12.5. Increased FGF signaling activation in TEPCs in *Sprouty* mutant mice has been shown to result in delayed initiation of *Foxn1* expression in the developing thymic primordium (Gardiner et al. 2012), suggesting that the observed downregulation of *Fgf8* expression is crucial for normal development.

3.2.2.3 *Pth* and *Gata3*: genes important in parathyroid development and function

Given that *Plet1* expression during development is not restricted to 3PP and the parathyroid domain within the 3PP also expresses *Plet1*, I tested for expression of *parathyroid hormone* (*Pth)* gene to determine whether the sorted PLET1^{+} cells may also contain parathyroid cells. *Pth* expression was not detected at E9.5, E10.5, and E11.0, after which the expression increased sharply till E12.5 (Figure 3.3E). The observed *Pth* expression is indicative of presence of parathyroid cells in the sorted population. On the other hand, analysis of the expression of the parathyroid specific marker, *Gcm2*, showed an absence of *Gcm2* expression in E12.5 samples (Figure 3.3F). *Gcm2* expression was detectable from E10.5 and increased by around 2-fold at E11.0 (p-value = 0.042), after which it remained constant till E12.0. The development of 3PP and subsequent expression of *Gcm2* in the parathyroid domain requires expression of *Gata3* within the cells of the common primordium (Grigorieva et al. 2010). *Gata3* is also important for postnatal thymus homeostasis (Blackburn lab unpublished) and thymocyte maturation (Ting et al. 1996; Hendriks et al. 1999).

Gata3 expression was detectable from as early as E9.5 and continued until E12.5 (Figure 3.3G). However, *Gata3* expression decreased significantly from E12.0 to E12.5 (p-value = 0.0006) indicating dynamic regulation of this gene beyond E12.0. Since GATA3 is required for *Gcm2* expression, one possibility is that its expression is down-regulated in TECs to prevent ectopic expression of parathyroid genes or to allow up-regulation of thymus genes. Together, these results suggest the presence of parathyroid cells within the analysed samples, however, the proportion of parathyroid cells within each sample cannot be determined based on the flow cytometry and gene expression data presented here.

3.2.2.4 *Notch1*, *Hes1*, and *Hes6*: genes involved in Notch-signaling

Notch signaling is critical for commitment of ETPs to the T-lymphocyte lineage, however its role in TEC biology remains less well characterized (Felli et al. 1999; Masuda et al. 2009; Lmmerts van Beuren et al. 2010; Shakib et al. 2009). Interestingly, the expression of *Notch1* was significantly up-regulated from E12.0 to E12.5 (p-value = 0.029) (Figure 3.3H), suggesting that Notch signaling pathway may play an important role in thymus development beyond E12.0. The up-regulation of *Notch1* at E12.5, a time point with highest expression of *Foxn1*, is consistent with the regulation of *Notch1* expression in hair follicle by FOXN1 (Cai et al. 2009). Thus, these results suggest that the expression of *Notch1* in TECs might also be regulated by FOXN1. The expression of *Hes1*, one of the key target genes of the Notch signaling pathway in neural cells, was detected from E9.5 and increased gradually till E12.5 (Figure 3.3I). HES6, a negative regulator of Notch signaling, has been shown to suppress the transcriptional activity of HES1 during neurogenesis, thus promoting neuronal differentiation (Bae et al. 2000; Koyano-Nakagawa et al. 2000). Biological replicates showed considerable variation in *Hes6* expression at E9.5 making it difficult to determine whether or not this gene was expressed at this developmental time point (Figure 3.3J). However, *Hes6* expression was detected from E10.5 and showed generally increasing expression until E12.5, except at E11.5 when its expression dropped transiently. Of note, the expression levels of *Hes6* are

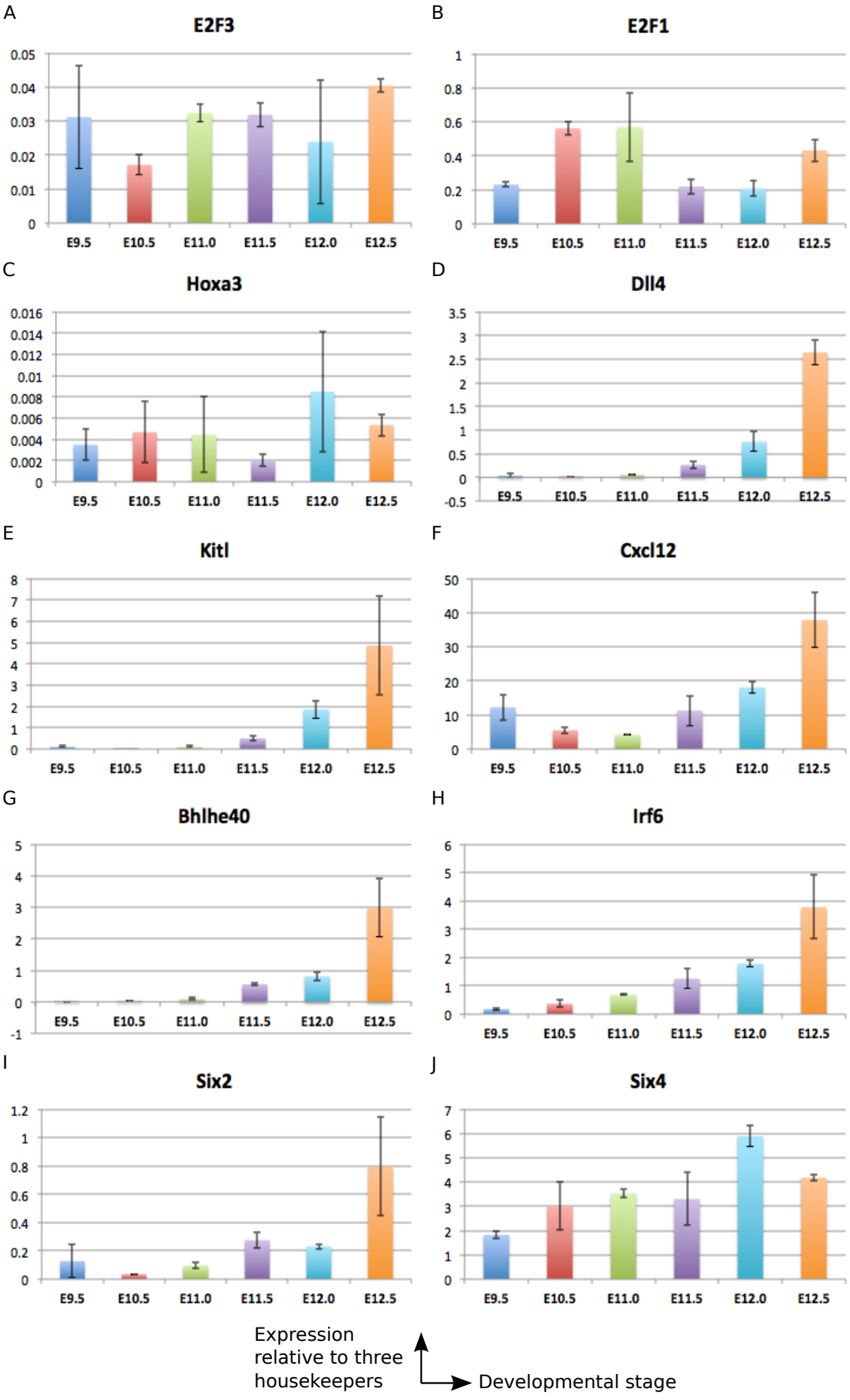
several order of magnitude lower that that of *Hes1*. Thus, it is likely that Notch signaling is important in thymus development.

3.2.2.5 Other genes

It has recently been shown that E2F3 and E2F4 can regulate *Foxn1* expression in postnatal TECs (Garfin et al. 2013). The E2F transcription factors play an important role in the control of cell cycle progression, via binding to the Rb family of proteins. However, the phenotypic difference in the mutant mice used in this study is only evident postnatally, suggesting that E2F3 and E2F4 mediated regulation of *Foxn1* expression may not be important in TEPCs of the developing thymus. The expression of *E2F3* was unchanged in TEPCs across the developmental time points analyzed (Figure 3.4A). Interestingly, the authors suggested a requirement of FOXN1 for up-regulation of *E2F1* expression in mutant TECs (Garfin et al. 2013). Analyzing the expression of *E2F1* in developing TEPCs revealed an interesting pattern with a modest decrease in expression after E11.0 (Figure 3.4B). Subsequently, the expression of *E2F1* increased modestly from E11.5 to E12.5, concomitant with up-regulation of *Foxn1* expression. While the expression patterns for E2F3 and E2F1 do not suggest involvement in regulating initiation of *Foxn1* expression, it is possible these transcription factors are still involved in regulating *Foxn1* expression as their activities are regulated through post-translational by Rb proteins.

Figure 3.4: Analysis of *E2F3*, *E2F1*, *Hoxa3*, *Dll4*, *Kitl*, *Cxcl12*, *Bhlhe40*, *Irf6*, *Six2*, *and Six4* expression patterns during normal thymus organogenesis.

See following page
Relative expression levels in 3PP and TEP cells were determined by QRT-PCR. Graphs show the expression patterns in WT thymus for (A) *E2F3*, (B) *E2F1*, (C) *Hoxa3*, (D) *Dll4*, (E) *Kitl*, (F) *Cxcl12*, (G) *Bhlhe40*, (H) *Irf6*, (I) *Six2*, and (J) *Six4*. Data are shown relative to the geometric mean Ct value for three housekeeping genes (*HMBS*, *b-actin*, *TBP*). Data shown are representative of two biological replicates and three technical replicates. Error bars show standard deviation (SD).


A
E2F3
0.05
0.04
0.03
0.02
0.01
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
B
E2F1
1
0.8
0.6
0.4
0.2
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
C
Hoxa3
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
D
Dll4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
E
Kitl
8
7
6
5
4
3
2
1
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
F
Cxcl12
50
40
30
20
10
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
G
Bhlhe40
5
4
3
2
1
0
-1
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
H
Irf6
6
5
4
3
2
1
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
I
Six2
1.2
1
0.8
0.6
0.4
0.2
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
J
Six4
7
6
5
4
3
2
1
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
Expression relative to three housekeepers
Developmental stage

Hoxa3, a gene important for 3PP and thymus development (Manley & Capecchi 1995; Manley & Capecchi 1998; Su et al. 2001), was expressed at relatively constant but very low levels from E9.5 to E12.5 (Figure 3.4C). This is consistent with it not being required for initiation of *Foxn1* expression (Chojnowski et al. 2014). As mentioned above, *Dll4*, *Cxcl12*, and *Kitl* are thought to be direct targets of FOXN1. Consistent with this, *Dll4* and *Kitl* expression was not detected prior to E11.0 and their expression increased from E11.0 to E12.5 (Figure 3.4D & Figure3.4E). The expression of *Cxcl12* was detectable from E9.5, before the initiation of *Foxn1* expression, and increased from E11.5 to E12.5 (p-value = 0.046 (E11.5 vs E12.5)) (Figure 3.4F). This suggests that its expression prior to E11.5 might be *Foxn1* independent. *Bhlhe40* (Figure 3.4G) also showed expression pattern similar to FOXN1 target genes, suggesting it too might be regulated by FOXN1. The expression of *Irf6* was similar to that expected of a positive transcriptional regulator of *Foxn1* (Figure 3.4H).

Six2 was expressed at low levels till E12.0, however its expression was sharply (but not significantly) up-regulated at E12.5 (Figure 3.4I), similar to the observation for *Notch1*. Another gene belonging to the *sine oculis-related homeobox* family, *Six4* was expressed at much higher levels than *Six2*. *Six4* expression increased slightly from E9.5 to E10.5 and peaked at E12.0 (Figure 3.4J). *Six4* has previously been suggested to be important for organ-specific gene regulation in 3PP (Xu et al. 2002; Zou et al. 2006), however its expression during thymus development beyond E10.5 was previously not studied. The expression pattern observed here suggests that *Six4* is unlikely to be involved in regulation of *Foxn1* expression.

It has previously been shown that *Eya1* is required for normal development of the thymus and parathyroid from 3PP (Xu et al. 2002; Zou et al. 2006). *Eya1* expression increased significantly from E9.5 to E10.5 (p-value = 0.014) and then decreased after E12.0 to a level similar to that seen at E9.5 (Figure 3.5A). Thus, expression of *Eya1* seemed to be anti-correlated with that of *Foxn1*, suggesting a possible negative regulation of *Eya1* by *Foxn1*. *Foxc1*, a gene important for tumor-suppressor

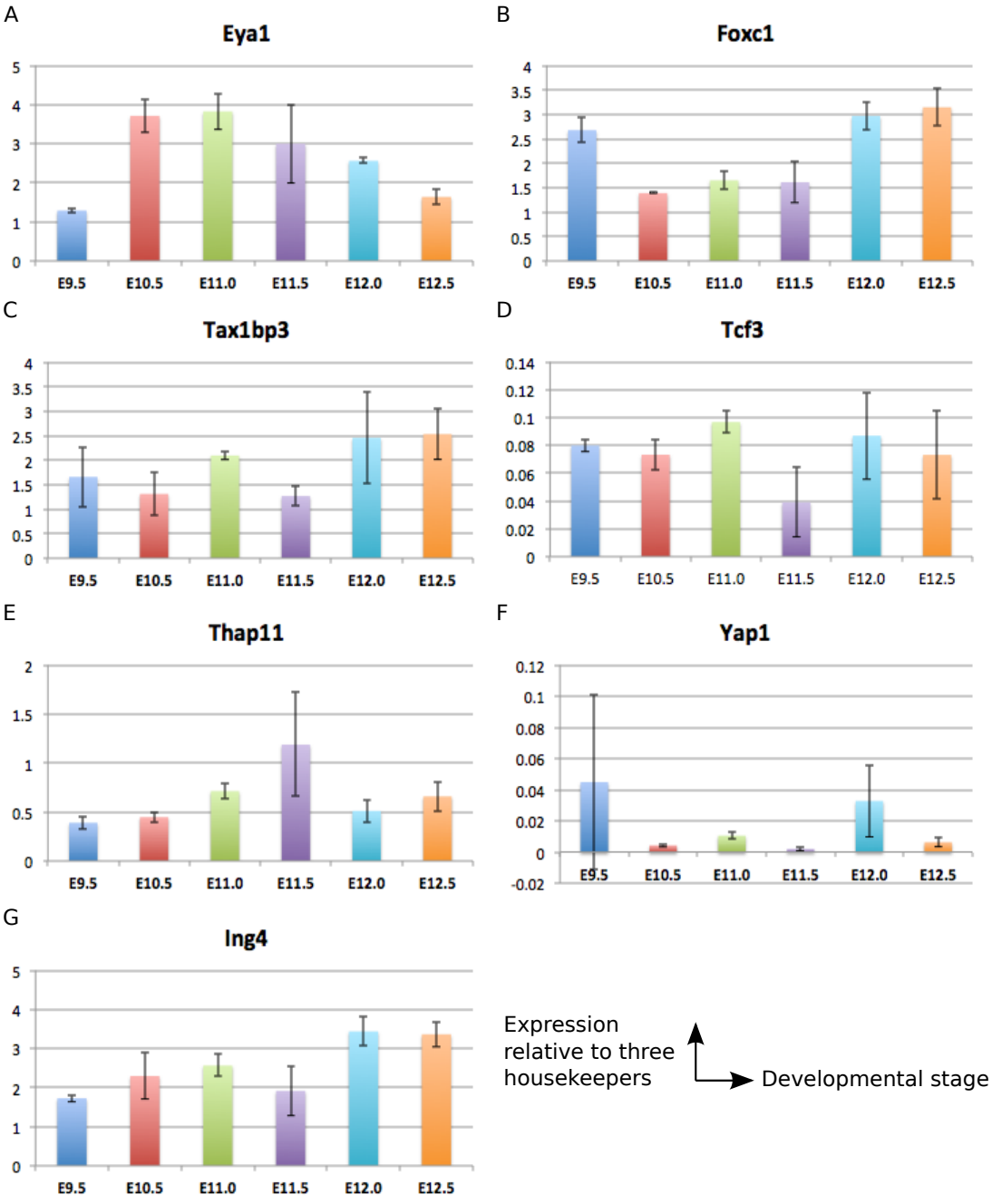


Figure 3.5: Analysis of *Eya1*, *Foxc1*, *Tax1bp3*, *Tcf3*, *Thap11*, *Yap1*, and *Ing4* expression profiles during normal thymus organogenesis.

Relative expression levels in 3PP and TEP cells were determined by QRT-PCR. Graphs show the expression profiles in WT thymus for (A) *Eya1*, (B) *Foxc1*, (C) *Tax1bp3*, (D) *Tcf3*, (E) *Thap11*, (F) *Yap1*, and (G) *Ing4*. Data are shown relative to the geometric mean Ct value for three housekeeping genes (*HMBS*, *b-actin*, *TBP*). Data shown are representative of two biological replicates and three technical replicates. Error bars show standard deviation (SD).

activities and prognosis of thymic epithelial tumor progression, was expressed from E9.5 and its expression was lower from E10.5 to E11.5 as compared to other time points analyzed (Figure 3.5B). Other genes, such as *Tax1 (human T cell leukemia virus type I) binding protein 3* (*Tax1bp3*; Figure 3.5C), *Transcription factor 3* (*Tcf3*; Figure 3.5D), *THAP domain containing 11* (*Thap11*; Figure 3.5E), and *Yes-associated protein 1* (*Yap1*; Figure 3.5F) showed no or little change in expression between the time points analyzed, suggesting that these genes are unlikely to be involved in regulating *Foxn1* expression. The expression of *inhibitor of growth family, member-4* (*Ing4)* – a tumor suppressor gene important for prostate epithelial cell differentiation – remained constant from E9.5 to E11.5 after which it increased slightly at E12.0 (Figure 3.5G). Whether the observed upregulation is important for *Foxn1* expression or is a result of increased FOXN1 levels remains to be determined.

Together, the above gene expression analysis identified several genes whose expression is similar to an expected expression pattern for positive transcriptional regulator of *Foxn1*. On the other hand, it also identified genes whose expression pattern suggest that they are unlikely to be involved in this process.

3.2.3 Comparison of gene expression between WT and *Foxn1*$^{-/-}$ TEPCs.

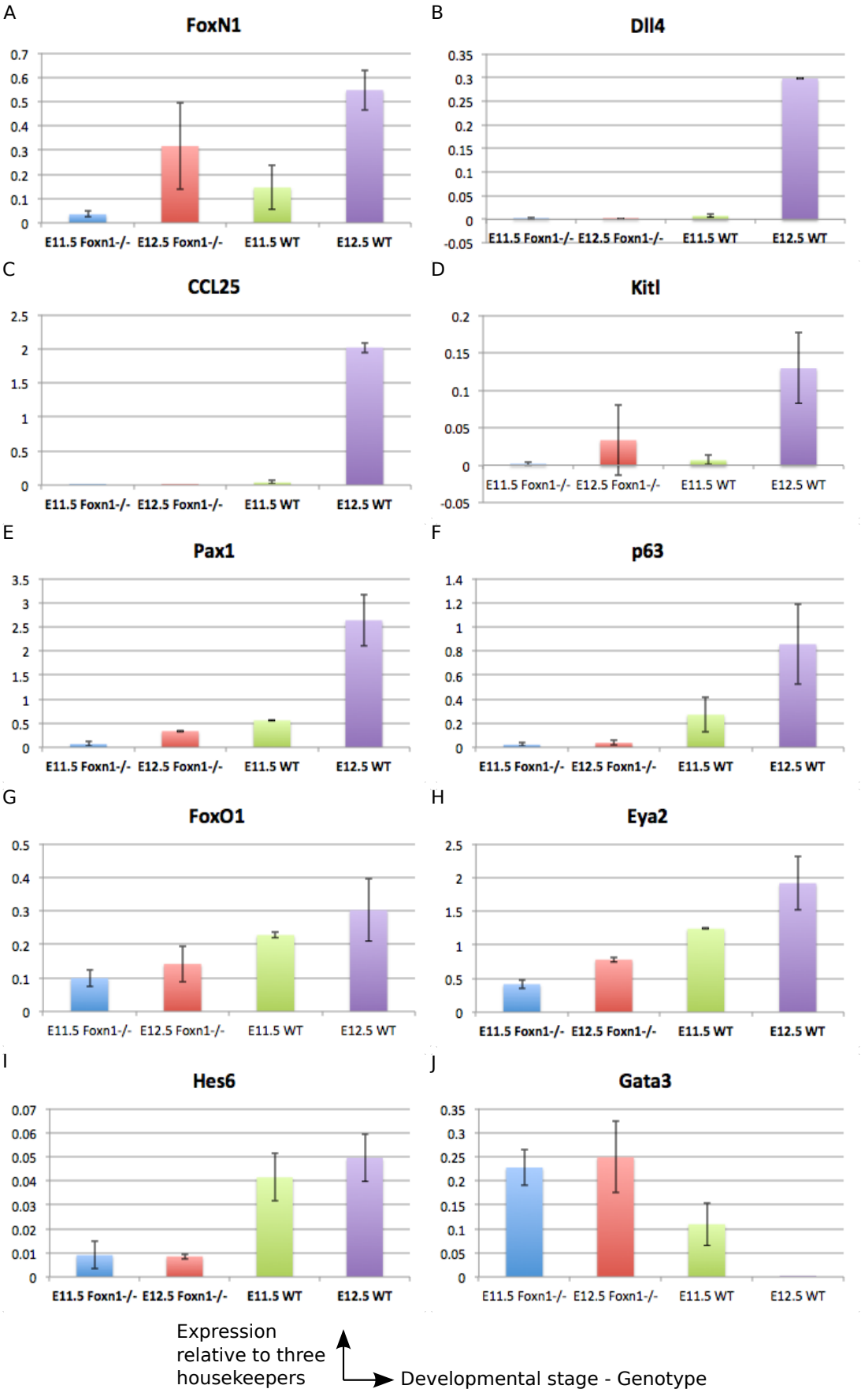
The above analysis revealed several interesting patterns. Most of the genes analyzed showed at least some degree of expression before the onset of *Foxn1* expression. While genes such as *Hoxa3, Pax1, Pax9, Bmp4, Fgf8, Eya1,* and *Six1* have been shown to be important for 3PP formation, the detailed expression of these genes, except *Hox3*, *Pax1*, and *Pax9*, during thymus development was not previously studied. Thus, while identifying new genes and demonstrating their expression patterns during thymus development, the above data also provide novel insights into the expression of genes with known thymus phenotypes. However, the above approach fails to distinguish between transcriptional regulators of *Foxn1* and transcriptional targets of FOXN1. For example, the increase in expression of genes from E11.0 to E12.5 could either result from an increase in the expression of *Foxn1*,

or vice versa. Also, transcriptional regulatory networks often contain feedback and feed-forward loops, which further complicates this issue. The presence of genes such as *Dll4*, *Kitl*, and *Cxcl12*, whose expressions is FOXN1 dependent, in the list of genes from my initial correlation analysis exemplifies this issue. Thus, it was important to determine whether the expression of any of the proposed transcriptional regulators of *Foxn1* in TEPCs were FOXN1-dependent. To do this, I sorted $EpCAM^{+}PLET1^{+}$ TEPCs from WT and $Foxn1^{-/-}$ embryos at two developmental stages: E11.5, the time of initiation of *Foxn1* expression, and E12.5, when most of the cells in thymic primordium express high levels of *Foxn1*. These time points also allow comparison between results from this analysis and those from the above time series analysis, thus improving the confidence in the observed results.

The *Foxn1* allele used for generating $Foxn1^{-/-}$ embryos was a knock-in, knock-out allele which disrupts exon3 of *Foxn1* with an IRES-LacZ cassette (Nehls et al. 1996). Thus, as would be expected since the probe was located in exon2, *Foxn1* expression was detectable in both WT and $Foxn1^{-/-}$ TEPCs (Figure 3.6A). The expression of *Foxn1* was slightly lower in $Foxn1^{-/-}$ than WT TEPC, consistent with the proposed positive auto-regulation of *Foxn1* by FOXN1 protein (Zook et al. 2011; Bredenkamp, Nowell, et al. 2014). Of note, a 10-fold decrease in the expression level was observed for all genes analysed here, compared to their expression in the previous analysis during thymus development, however the reason for this is unclear. One possibility is that the

Figure 3.6: Analysis of *Foxn1*, *Dll4*, *Ccl25*, *Kitl*, *Pax1*, *p63*, *Foxo1*, *Eya2*, *Hes6*, *and Gata3* expression patterns in E11.5 and E12.5 TEPCs isolated from wild type and *Foxn1* null thymi.

See following page
Relative expression levels in wild type and *Foxn1* null E11.5 and E12.5 TEPCs were determined by QRT-PCR. Graphs show the expression patterns for (A) *Foxn1*, (B) *Dll4*, (C) *Ccl25*, (D) *Kitl*, (E) *Pax1*, (F) *p63*, (G) *Foxo1*, (H) *Eya2*, (I) *Hes6*, and (J) *Gata3*. Data are shown relative to the geometric mean Ct value for three housekeeping genes (*HMBS*, *b-actin*, *TBP*). Data shown are representative of two biological replicates and three technical replicates. Error bars show standard deviation (SD).


A
FoxN1
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
B
Dll4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
C
CCL25
2.5
2
1.5
1
0.5
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
D
Kitl
0.2
0.15
0.1
0.05
0
-0.05
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
E
Pax1
3.5
3
2.5
2
1.5
1
0.5
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
F
p63
1.4
1.2
1
0.8
0.6
0.4
0.2
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
G
FoxO1
0.5
0.4
0.3
0.2
0.1
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
H
Eya2
2.5
2
1.5
1
0.5
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
I
Hes6
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
J
Gata3
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
Expression relative to three housekeepers
Developmental stage - Genotype

The expression of *Dll4* (Figure 3.6B) and *Ccl25* (Figure 3.6C) was significantly lower in *Foxn1*$^{-/-}$ TEPCs compared to WT (p-value < 0.05). This is in accordance with proposed transcriptional regulation of *Dll4* and *Ccl25* by FOXN1. A similar pattern was observed for *Kitl* expression (Figure 3.6D), however substantial variation between two biological replicates of E12.5 *Foxn1*$^{-/-}$ TEPCs was observed. Surprisingly, there was an 8-fold reduction in the expression of *Pax1* in *Foxn1*$^{-/-}$ TEPCs at both E11.5 and E12.5, compared to age-matched WT TEPCs (p-value = 0.06 (E11.5); p-value = 0.018 (E12.5)) (Figure 3.6E), suggesting that FOXN1 reinforces the expression of *Pax1* in TEPCs through feedback mechanism.

As described above, the expression of *p63* increased from E11.5 to E12.5. However, this increase in expression was not evident in *Foxn1*$^{-/-}$ TEPCs (Figure 3.6F). In fact, the expression of *p63* at both E11.5 and E12.5 was significantly lower in *Foxn1*$^{-/-}$ compared to WT TEPCs (p-value = 0.04 (E11.5); p-value = 0.015 (E12.5)). This suggest that *p63* is downstream of *Foxn1* in the transcriptional network prevalent in TEPCs. Other genes that showed a similar pattern of reduced expression in *Foxn1*$^{-/-}$ vs WT TEPCs include *Foxo1* (p-value = 0.004) (Figure 3.6G), *Eya2* (p-value = 0.003) (Figure 3.6H), and *Hes6* (p-value = 0.0004) (Figure 3.6I). Of interest is that the expression of *Hes6*, a negative regulator of *Hes1*, appears to be FOXN1-dependant, suggesting a role of *Foxn1* in regulating the downstream Notch pathway. On the other hand, the expression of *Gata3* decreased from E11.5 to E12.5 in WT TEPCs, with no detectable expression at the later time point, but its expression remained unchanged in *Foxn1*$^{-/-}$ TEPCs (Figure 3.6J) suggesting that FOXN1 represses *Gata3* expression in these cells. The expression profiles of the above-mentioned genes were significantly correlated with each other (or anti-correlated in the case of *Gata3*) as evident from gene-wise correlation coefficient analysis carried out using the R package, Hmisc (details below).

Other genes that showed a decrease in expression from E11.5 to E12.5 in WT but not *Foxn1*$^{-/-}$ TEPCs include *Eya1* (Figure 3.7A), *Eid1* (Figure 3.7B), *p53* (Figure 3.7C),

Tax1bp3 (Figure 3.7D), and *Thap11* (Figure 3.7E). The decrease in expression of *Eya1* and *Thap11* from E11.5 to E12.5 WT TEPCs is consistent with the previously described expression profile during thymus development and appears to be dependent on the presence of a functional FOXN1 protein. Thus, these genes are likely to be regulated by FOXN1 in TEPCs.

Interestingly, *Foxc1* (Figure 3.8A) was expressed more highly in WT than *Foxn1*$^{-/-}$ TEPCs. The role of these Forkhead genes in thymus development is not yet known, however these results suggest that FOXN1 could play a role in regulating the expression of *Foxc1* in TEPCs. Another member of the Forkhead gene family, *Foxg1*, showed an increase in expression from E11.5 to E12.5 in *Foxn1*$^{-/-}$ TEPCs while its expression remained unchanged in WT TEPCs, suggesting possible compensation mechanism between the two genes (Figure 3.8B). Thus, these results indicate a complex regulatory network involving the various *Fox* genes. Further experiments involving knock-out/knock-down of each of these genes are required to better understand these relationships. Similar to the developmental time point analysis, considerable variation was observed in the expression of *Foxa2* (Figure 3.8C). However, its expression decreased from E11.5 to E12.5 in both WT and *Fxon1*$^{-/-}$ cells, suggesting that the down-regulation of its expression during development is *Foxn1* independent. As mentioned previously, *Six1* and *Six4* play an important role in 3PP development and thymus organogenesis. The expression of both these genes was higher at E11.5 in WT TEPCs as compared to *Foxn1*$^{-/-}$ TEPCs (p-value = 0.036 (Six1); p-value = 0.047 (Six4)) (Figure 3.8D and Figure 3.8E), however no difference in expression was observed at E12.5. The expression of *Fgf8* was detectable at E11.5 but not at E12.5 in both WT and *Foxn1*$^{-/-}$ TEPCs (Figure 3.8F), consistent with its regulation by TBX1 (Vitelli et al. 2002; Huh & Ornitz 2010; Guo et al. 2011).

Bhlhe40, a member of basic helix-loop-helix transcriptional regulators, which are known to respond to environmental stimuli and regulate several biological processes

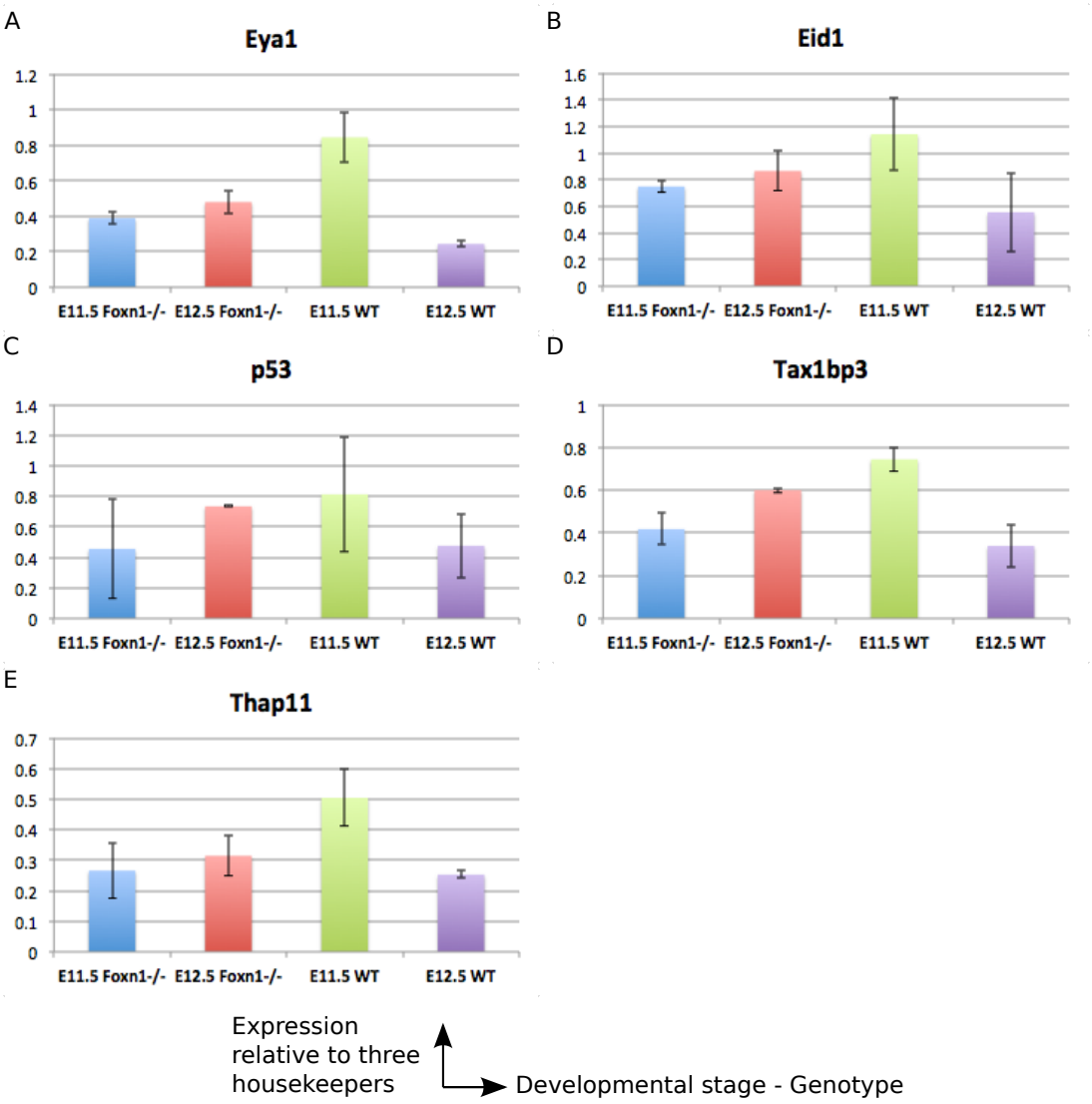


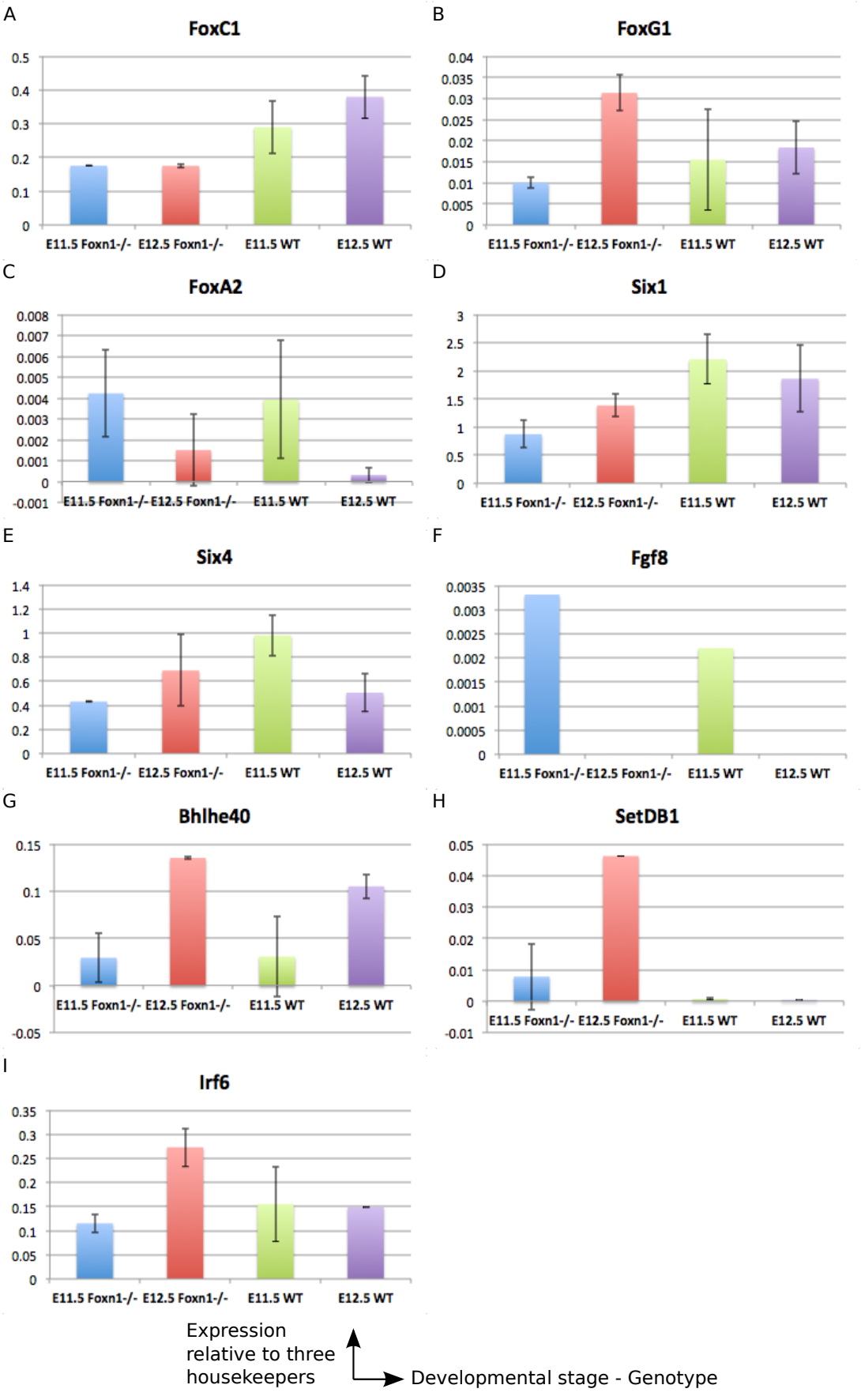
Figure 3.7: Analysis of *Eya1*, *Eid1*, *p53*, *Tax1bp3*, and *Thap11* expression profiles in E11.5 and E12.5 TEPCs isolated from wild type and *Foxn1* null thymi.
Relative expression levels in wild type and *Foxn1* null E11.5 and E12.5 TEPCs were determined by QRT-PCR. Graphs show the expression profiles in WT thymus for (A) *Eya1*, (B) *Eid1*, (C) *p53*, (D) *Tax1bp3*, and (E) *Thap11*. Data are shown relative to the geometric mean Ct value for three housekeeping genes (*HMBS*, *b-actin*, *TBP*). Data shown are representative of two biological replicates and three technical replicates. Error bars show standard deviation (SD).

by acting as both transcriptional activators and repressors, has been shown to be important for normal T cell function (Lin et al. 2014). However, its role in TECs is not yet known. Interestingly, the expression of *Bhlhe40* increased from E11.5 to E12.5 independently of *Foxn1* (Figure 3.8G). Such an expression pattern would be expected for a transcriptional regulator of *Foxn1*, in the absence of any immediate feedback mechanism, suggesting *Bhlhe40* could be involved in regulating *Foxn1* expression. The expression of *SetDB1*, a gene encoding a H3K9 methyltransferase was detectable in *Foxn1*$^{-/-}$ TEPCs but not in WT TEPCs (Figure 3.8H), suggesting a potential role of *Foxn1* in regulating epigenetic modification in developing TECs. Contrary to the previous observation, the expression of *Irf6* did not change between E11.5 and E12.5 WT TEPCs (Figure 3.8I), suggesting variability in analyzing its expression. On the other hand, the increase in *Irf6* expression at E12.5 compared to E11.5 in *Foxn1*$^{-/-}$ TEPCs is consistent with the previous observation (Figure 3.4H). Thus, its expression in TEPCs appears to be independent of *Foxn1*.

The final class of genes in this analysis was the genes whose expression did not change substantially between the analyzed samples. Thus, these genes appear to be insensitive (or less sensitive) to loss of functional FOXN1 protein, however their role in regulating *Foxn1* expression cannot be excluded based on current analysis alone. These genes include *Zfp36l1* (Figure 3.9A), *Hey1* (Figure 3.9B), *Eif3a* (Figure 3.9C), *Fbxw7* (Figure 3.9D), *Creb3l2* (Figure 3.9E), *Zfp503* (Figure 3.9F), *Hoxa3* (Figure 3.9G), *Ing4* (Figure 3.9H), and *NfiB* (Figure 3.9I).

Figure 3.8: Analysis of *Foxc1*, *Foxg1*, *Foxa2*, *Six1*, *Six4*, *Fgf8*, *Bhlhe40*, *SetDB1*, *and Irf6* expression patterns in E11.5 and E12.5 TEPCs isolated from wild type and *Foxn1* null thymi.

See following page
Relative expression levels in wild type and *Foxn1* null E11.5 and E12.5 TEPCs were determined by QRT-PCR. Graphs show the expression patterns for (A) *Foxc1*, (B) *Foxg1*, (C) *Foxa2*, (D) *Six1*, (E) *Six4*, (F) *Fgf8*, (G) *Bhlhe40*, (H) *SetDB1*, and (J) *Irf6*. Data are shown relative to the geometric mean Ct value for three housekeeping genes (*HMBS*, *b-actin*, *TBP*). Data shown are representative of two biological replicates and three technical replicates. Error bars show standard deviation (SD).


A
FoxC1
0.5
0.4
0.3
0.2
0.1
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
B
FoxG1
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
C
FoxA2
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
-0.001
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
D
Six1
3
2.5
2
1.5
1
0.5
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
E
Six4
1.4
1.2
1
0.8
0.6
0.4
0.2
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
F
Fgf8
0.0035
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
G
Bhlhe40
0.15
0.1
0.05
0
-0.05
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
H
SetDB1
0.05
0.04
0.03
0.02
0.01
0
-0.01
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
I
Irf6
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
Expression relative to three housekeepers
Developmental stage - Genotype

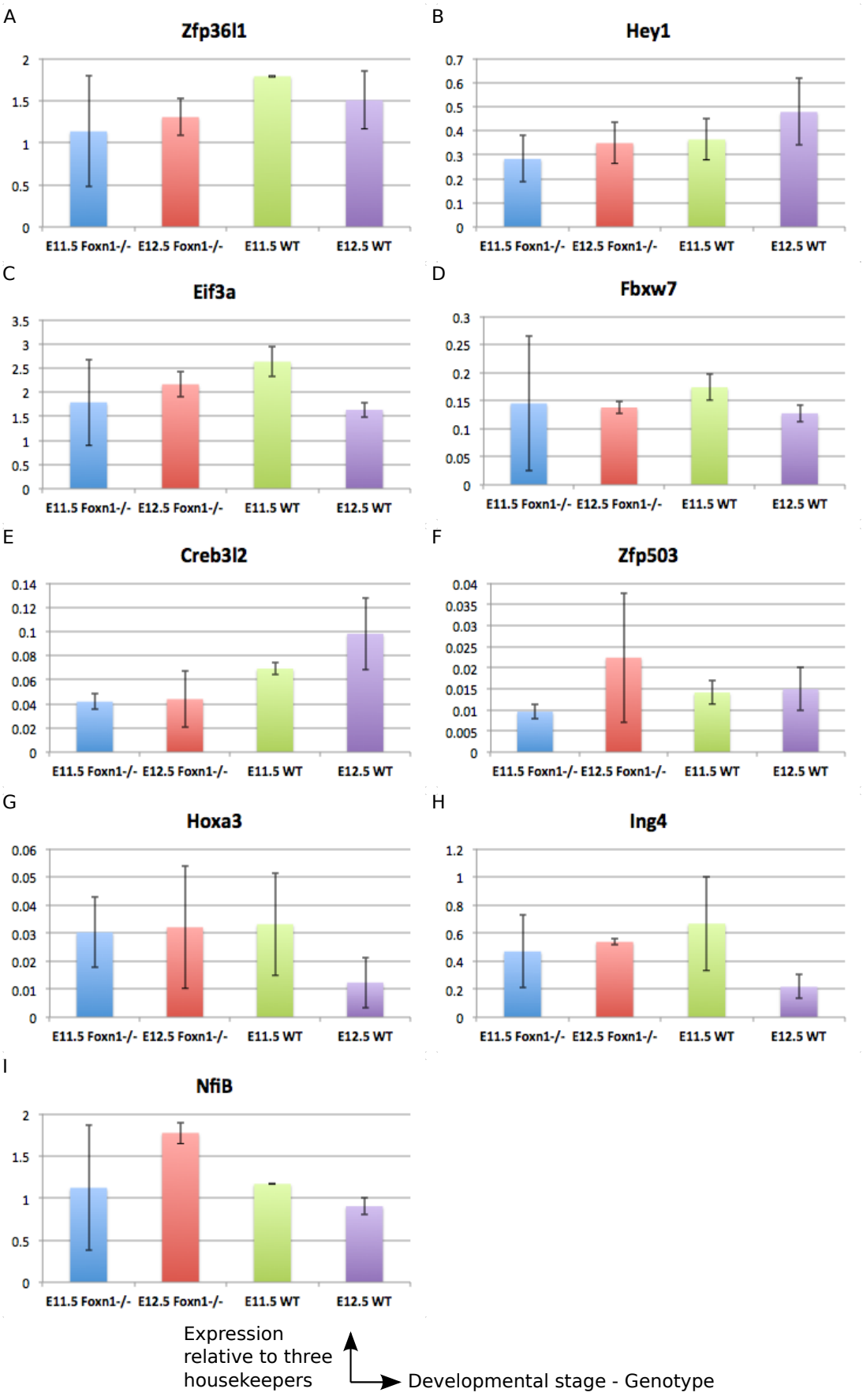
3.2.3.1 Gene-wise correlation analysis to clusters of genes with significantly correlated expression.

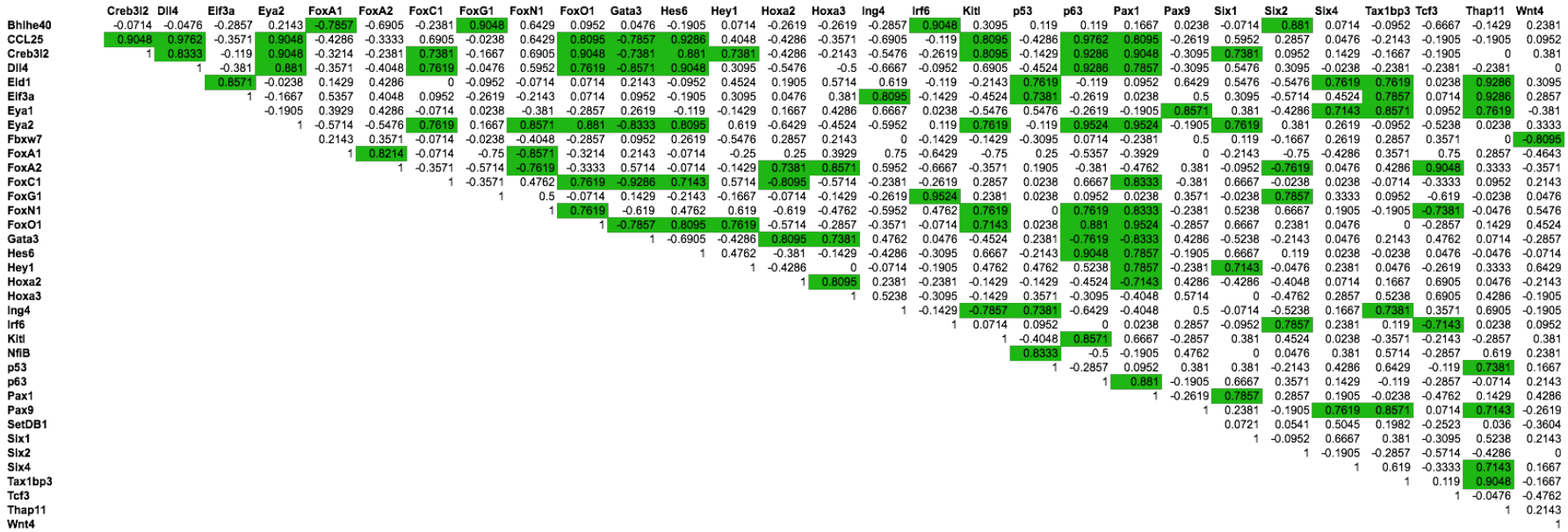
The above analysis of WT and *Foxn1*$^{-/-}$ TEPCs generated expression profiles for more than 40 genes at two developmental time points in both WT and *Foxn1*$^{-/-}$ TEPCs. These profiles can be used to infer complex patterns of gene regulatory network prevalent in the cells. To do this, I used Hmisc, a R-package for data analysis, to generate gene-wise correlation coefficient values and corresponding *p-values* to determine the statistical significance of each correlation coefficient. Figure 3.10 shows the correlation coefficient for all gene pairs, with statistically significant coefficients being highlighted in green. As noted above, the *Foxn1*$^{-}$ allele is transcribed and the transcript detected by the primer pairs used in this analysis. Thus, the correlation coefficients generated using *Foxn1* do not truly represent the conditions *in-vivo*, which hinders identification of genes within in the same transcriptional network as *Foxn1*. To partly overcome this issue, I decided to use the expression of two *Foxn1*-dependant genes, *Ccl25* and *Dll4* as they resemble the expected expression pattern for *Foxn1* as shown above. This approach could help identify other genes whose expression is also dependent on FOXN1. Figure 3.11 shows genes whose expression is significantly correlate or anti-correlated with that

Figure 3.9: Analysis of *Zfp36l1*, *Hey1*, *Eif3a*, *Fbxw7*, *Creb3l2*, *Zfp503*, *Hoxa3*, *Ing4*, *and NfiB* expression patterns in E11.5 and E12.5 TEPCs isolated from wild type and *Foxn1* null thymi.

See following page

Relative expression levels in wild type and *Foxn1* null E11.5 and E12.5 TEPCs were determined by QRT-PCR. Graphs show the expression patterns for (A) *Zfp36l1*, (B) *Hey1*, (C) *Eif3a*, (D) *Fbxw7*, (E) *Creb3l2*, (F) *Zfp503*, (G) *Hoxa3*, (H) *Ing4*, and (J) *NfiB*. Data are shown relative to the geometric mean Ct value for three housekeeping genes (*HMBS*, *b-actin*, *TBP*). Data shown are representative of two biological replicates and three technical replicates. Error bars show standard deviation (SD).


A
Zfp36l1
2
1.5
1
0.5
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
B
Hey1
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
C
Eif3a
3.5
3
2.5
2
1.5
1
0.5
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
D
Fbxw7
0.3
0.25
0.2
0.15
0.1
0.05
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
E
Creb3l2
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
F
Zfp503
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
G
Hoxa3
0.06
0.05
0.04
0.03
0.02
0.01
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
H
Ing4
1.2
1
0.8
0.6
0.4
0.2
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
I
NfiB
2
1.5
1
0.5
0
E11.5 Foxn1-/-
E12.5 Foxn1-/-
E11.5 WT
E12.5 WT
Expression relative to three housekeepers
Developmental stage - Genotype



	Creb3l2	Dll4	Elf3a	Eya2	FoxA1	FoxA2	FoxC1	FoxG1	FoxN1	FoxO1	Gata3	Hes6	Hey1	Hoxa2	Hoxa3	Ing4	Irf6	Kitl	p53	p63	Pax1	Pax9	Six1	Six2	Six4	Tax1bp3	Tcf3	Thap11	Wnt4
Bhlhe40	-0.0714	-0.0476	-0.2857	0.2143	-0.7857	-0.6905	-0.2381	0.9048	0.6429	0.0952	0.0476	-0.1905	0.0714	-0.2619	-0.2619	-0.2857	0.9048	0.3095	0.119	0.119	0.1667	0.0238	-0.0714	0.881	0.0714	-0.0952	-0.6667	-0.1429	0.2381
CCL25	0.9048	0.9762	-0.3571	0.9048	-0.4286	-0.3333	0.6905	-0.0238	0.6429	0.8095	-0.7857	0.9286	0.4048	-0.4286	-0.3571	-0.6905	-0.119	0.8095	-0.4286	0.9762	0.8095	-0.2619	0.5952	0.2857	0.0476	-0.2143	-0.1905	-0.1905	0.0952
Creb3l2	1	0.8333	-0.119	0.9048	-0.3214	-0.2381	0.7381	-0.1667	0.6905	0.9048	-0.7381	0.881	0.7381	-0.4286	-0.2143	-0.5476	-0.2619	0.8095	-0.1429	0.9286	0.9048	-0.3095	0.7381	0.0952	0.1429	-0.1667	-0.1905	0	0.381
Dll4		1	-0.381	0.881	-0.3571	-0.4048	0.7619	-0.0476	0.5952	0.7619	-0.8571	0.9048	0.3095	-0.5476	-0.5	-0.6667	-0.0952	0.6905	-0.4524	0.9286	0.7857	-0.3095	0.5476	0.3095	-0.0238	-0.2381	-0.2381	-0.2381	0
Eld1			0.8571	-0.0238	0.1429	0.4286	0	-0.0952	-0.0714	0.0714	0.2143	-0.0952	0.4524	0.1905	0.5714	0.619	-0.119	-0.2143	0.7619	-0.119	0.0952	0.6429	0.5476	-0.5476	0.7619	0.7619	0.0238	0.9286	0.3095
Elf3a			1	-0.1667	0.5357	0.4048	0.0952	-0.2619	-0.2143	0.0714	0.0952	-0.1905	0.3095	0.0476	0.381	0.8095	-0.1429	-0.4524	0.7381	-0.2619	0.0238	0.5	0.3095	-0.5714	0.4524	0.7857	0.0714	0.9286	0.2857
Eya1				-0.1905	0.3929	0.4286	-0.0714	0.0238	-0.381	-0.2857	0.2619	-0.119	-0.1429	0.1667	0.4286	0.6667	0.0238	-0.5476	0.5476	-0.2619	-0.1905	0.8571	0.381	-0.4286	0.7143	0.8571	0.0952	0.7619	-0.381
Eya2				1	-0.5714	-0.5476	0.7619	0.1667	0.8571	0.881	-0.8333	0.8095	0.619	-0.6429	-0.4524	-0.5952	0.119	0.7619	-0.119	0.9524	0.9524	-0.1905	0.7619	0.381	0.2619	-0.0952	-0.5238	0.0238	0.3333
Fbxw7					0.2143	0.3571	-0.0714	-0.0238	-0.4048	-0.2857	0.0952	0.2619	-0.5476	0.2857	0.2143	0	-0.1429	-0.1429	-0.3095	0.0714	-0.2381	0.5	0.119	-0.1667	0.2619	0.2857	0.3571	0	-0.8095
FoxA1					1	0.8214	-0.0714	-0.75	-0.8571	-0.3214	0.2143	-0.0714	-0.25	0.25	0.3929	0.75	-0.6429	-0.75	0.25	-0.5357	-0.3929	0	-0.2143	-0.75	-0.4286	0.3571	0.75	0.2857	-0.4643
FoxA2						1	-0.3571	-0.5714	-0.7619	-0.3333	0.5714	-0.0714	-0.1429	0.7381	0.8571	0.5952	-0.6667	-0.3571	0.1905	-0.381	-0.4762	0.381	-0.0952	-0.7619	0.0476	0.4286	0.9048	0.3333	-0.3571
FoxC1							1	-0.3571	0.4762	0.7619	-0.9286	0.7143	0.5714	-0.8095	-0.5714	-0.2381	-0.2619	0.2857	0.0238	0.6667	0.8333	-0.381	0.6667	-0.0238	0.0238	-0.0714	-0.3333	0.0952	0.2143
FoxG1								1	0.5	-0.0714	0.1429	-0.2143	-0.1667	-0.0714	-0.1429	-0.2619	0.9524	0.2381	0.0238	0.0952	0.0238	0.3571	-0.0238	0.7857	0.3333	0.0952	-0.619	-0.0238	0.0476
FoxN1									1	0.7619	-0.619	0.4762	0.619	-0.619	-0.4762	-0.5952	0.4762	0.7619	0	0.7619	0.8333	-0.2381	0.5238	0.6667	0.1905	-0.1905	-0.7381	-0.0476	0.5476
FoxO1										1	-0.7857	0.8095	0.7619	-0.5714	-0.2857	-0.3571	-0.0714	0.7143	0.0238	0.881	0.9524	-0.2857	0.6667	0.2381	0.0476	0	-0.2857	0.1429	0.4524
Gata3											1	-0.6905	-0.4286	0.8095	0.7381	0.4762	0.0476	-0.4524	0.2381	-0.7619	-0.8333	0.4286	-0.5238	-0.2143	0.0476	0.2143	0.4762	0.0714	-0.2857
Hes6												1	0.4762	-0.381	-0.1429	-0.4286	-0.3095	0.6667	-0.2143	0.9048	0.7857	-0.1905	0.6667	0.119	0.0238	-0.0238	0.0476	-0.0476	-0.0714
Hey1													1	-0.4286	0	-0.0714	-0.1905	0.4762	0.4762	0.5238	0.7857	-0.2381	0.7143	-0.0476	0.2381	0.0476	-0.2619	0.3333	0.6429
Hoxa2														1	0.8095	0.2381	-0.2381	-0.1429	-0.1429	-0.4524	-0.7143	0.4286	-0.4286	-0.4048	0.0714	0.1667	0.6905	0.0476	-0.2143
Hoxa3															1	0.5238	-0.3095	-0.1429	0.3571	-0.3095	-0.4048	0.5714	0	-0.4762	0.2857	0.5238	0.6905	0.4286	-0.1905
Ing4																1	-0.1429	-0.7857	0.7381	-0.6429	-0.4048	0.5	-0.0714	-0.5238	0.1667	0.7381	0.3571	0.6905	-0.1905
Irf6																	1	0.0714	0.0952	0	0.0238	0.2857	-0.0952	0.7857	0.2381	0.119	-0.7143	0.0238	0.0952
Kitl																		1	-0.4048	0.8571	0.6667	-0.2857	0.381	0.4524	0.0238	-0.3571	-0.2143	-0.2857	0.381
NfiB																			0.8333	-0.5	-0.1905	0.4762	0	0.0476	0.381	0.5714	-0.2857	0.619	0.2381
p53																			1	-0.2857	0.0952	0.381	0.381	-0.2143	0.4286	0.6429	-0.119	0.7381	0.1667
p63																				1	0.881	-0.1905	0.6667	0.3571	0.1429	-0.119	-0.2857	-0.0714	0.2143
Pax1																					1	-0.2619	0.7857	0.2857	0.1905	-0.0238	-0.4762	0.1429	0.4286
Pax9																						1	0.2381	-0.1905	0.7619	0.8571	0.0714	0.7143	-0.2619
SetDB1																							0.0721	0.0541	0.5045	0.1982	-0.2523	0.036	-0.3604
Six1																							1	-0.0952	0.6667	0.381	-0.3095	0.5238	0.2143
Six2																								1	-0.1905	-0.2857	-0.5714	-0.4286	0
Six4																									1	0.619	-0.3333	0.7143	0.1667
Tax1bp3																										1	0.119	0.9048	-0.1667
Tcf3																											1	-0.0476	-0.4762
Thap11																												1	0.2143
Wnt4																													1

Figure 3.10: Correlation coefficient between each pair of genes based on expression profiles during thymus organogenesis.

The correlation coefficient between expression profiles of each pair of genes were identified using the R package, Hmisc. Statistical significance (p-values) for each coefficient is calculated by performing 1000 itereations of calculating correlation coefficient for randomized data. Data randomization was performed using the bootstrap technique. Statistically significant coefficeints ($p < 0.05$) are highlighted in green.

Correlation coefficient	**Dll4**	**Eya2**	**FoxO1**	**Gata3**	**Hes6**	**Kitl**	**p63**	**Pax1**	**Foxc1**
CCL25	0.9762	0.9048	0.8095	-0.7857	0.9286	0.8095	0.9762	0.8095	0.6905
Dll4	1	0.881	0.7619	-0.8571	0.9048	0.6905	0.9286	0.7857	0.7619

Figure 3.11: Genes signigicantly correlated with Ccl25 or Dll4 during thymus organogenesis.
The correlation coefficient between expression profiles of each pair of genes were identified using the R package, Hmisc. Statistical significance (p-values) for each coefficient is calculated by performing 1000 itereations of calculating correlation coefficient for randomized data. Data randomization was performed using the bootstrap technique. Genes whose expression was significant coefficeints ($p < 0.05$) with that of Ccl25 or Dll4 are shown here.

of *Ccl25* or *Dll4*. Genes whose expression was significantly correlated to that of either *Ccl25* or *Dll4* include *Eya2*, *Foxo1*, *Hes6*, *Kitl*, *p63*, *Pax1*, and *Foxc1*. Interestingly, *Gata3* was the only gene whose expression is significantly negatively correlated with that of both *Ccl25* and *Dll4*. This suggests that these genes are targets of FOXN1 in TEPCs. Of note, the expression of genes, which are either significantly correlated (or anti-correlated) with *Ccl25* or *Dll4,* was also significantly correlated (or anti-correlated in the case of *Gata3*) with each other. This further supports the argument that these genes are a part of the same genetic network and their expression being regulated by a common transcriptional regulator (possibly *Foxn1*) in TEPCs.

3.3 Discussion

The data shown here constitute the first high-throughput gene expression analysis carried out in TEPCs. This approach has allowed identification of the dynamic changes in expression of a large number of genes during thymus development and is a significant advance from the previously published results, which include detection of presence of absence of transcripts using *in-situ* hybridization for a very small number of genes. Furthermore, the data from time-series analysis combined with that from comparing WT and *Foxn1*$^{-/-}$ TEPCs has allowed for identification of a potential genetic network in this cell type (Figure 3.12).

Several papers published since 2013 have identified candidate regulators of *Foxn1* expression using a combination of *in-vitro* and *in-vivo* approaches. *Tbx1*, a gene important for the formation of third pharyngeal pouch (Jerome & Papaioannou 2001), has recently been shown to be involved in regulating *Foxn1* expression in thymus (Reeh et al. 2014). Forced expression of *Tbx1* in thymus-fated domain of 3PP severely down-regulated the expression of *Foxn1*, indicating a role of *Tbx1* in antagonising thymus fate (Reeh et al. 2014). This is consistent with the observed expression profile for *Tbx1* described here, suggesting a strong down-regulation of this gene at the onset of *Foxn1* expression. Another interesting study showed that deletion of RB family of proteins results in increased E2F transcription factors

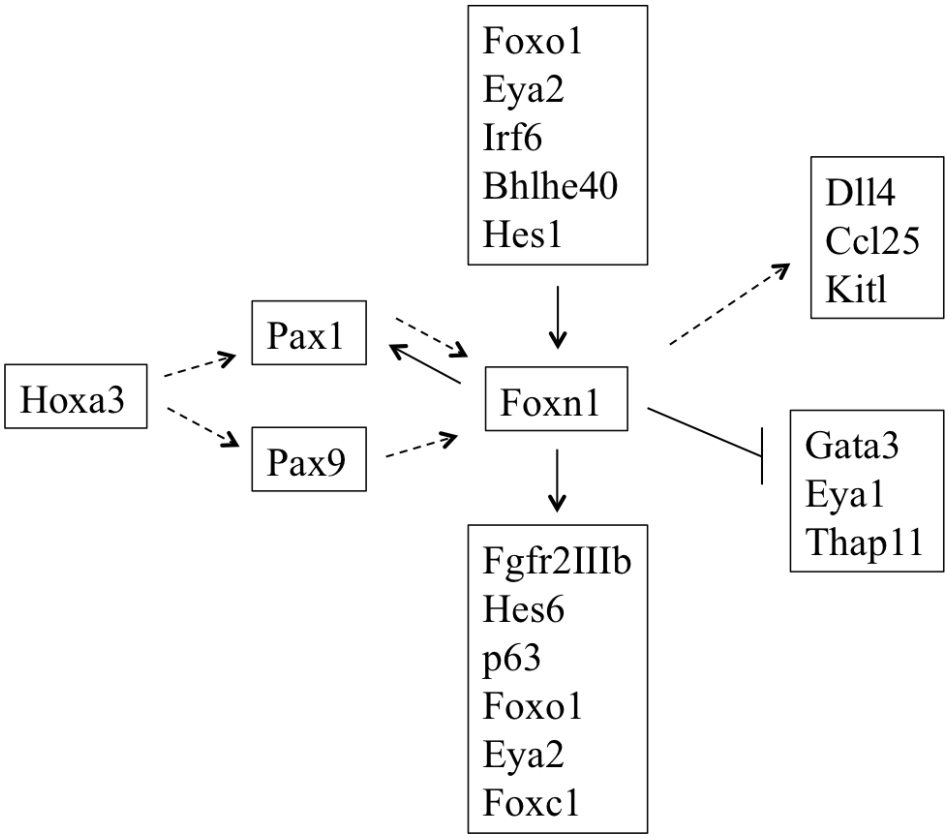


Figure 3.12: Predicted genetic network in TEPCs.
Genetic network predicted based on gene expression analysis of 3PP and TEPCs in WT and *nude* mice. Genes predicted to be transcriptional regulators of *Foxn1* or targets of FOXN1 protein are shown here. Also shown is regulation of *Pax1* by *Hoxa3* and that of *Foxn1* by *Pax1* known from previous studies. Interactions predicted from in this thesis are shown by solid lines whereas those known from literature are shown as dashed lines.

activity, leading to an enlarged thymus possibly resulting from increased expression of *Foxn1* in TECs (Garfin et al. 2013). This study further showed, using *in-vitro* approaches, that E2F3 and E2F4 could bind to conserved E2F binding sites near the transcriptional start site of *Foxn1* and thus regulate its expression. The expression of E2F3 did not change substantially across the developmental time points analysed herein. This could be because the activity of E2F proteins is regulated through post-translational regulation or it could indicate different regulatory mechanisms governing *Foxn1* during development and homeostasis, as the enlarged thymus phenotype in *Rb* mutant mice is not evident at 3 weeks of age. Of note is that the above mentioned studies lacked evidence of direct *in-vivo* regulation of *Foxn1* expression by any of these genes due to difficulties in performing a transcription factor ChIP using isolated TEPCs or TECs.

Pax1 and *Pax9*, genes shown to be important for normal thymus development, were recently identified as being important for maintenance (and possibly initiation) of *Foxn1* expression in the developing thymic primordium (Michelle Kelly, PhD thesis; Blackburn and Vaidya unpublished). The data shown here suggests that *Pax1* and *Foxn1* are members of a complex transcription factor network and that *Foxn1* may positively regulate *Pax1*, forming a feedback loop to increase the expression of *Pax1*. *Pax9* expression, on the other hand, does not appear to be dependent on FOXN1, suggesting that these two Pax genes are regulated differently. Furthermore, the data shown here are the first to suggest that FOXN1 may regulate *Gata3* expression. This is consistent with unpublished data in the lab showing that in adult thymus, the cells expression the highest levels of *Foxn1* have the lowest levels of *Gata3* expression and vice versa. Whether *Gata3* also plays a role in regulating *Foxn1* expression remains to be determined and would require analysing *Gata3* mutant thymi.

The presence of *Pth* expression in the analysed samples up to E12.0 indicates presence of parathyroid cells, which also express PLET1. However, it is not possible to determine the extent of contribution of parathyroid cells to the observed gene expression patterns in the above analyses. Identification of thymus specific cell-

surface marker or use of Gcm2-GFP alleles would be required to obtain samples free from parathyroid cells.

While a detailed analysis of the functions of FOXN1 in thymic epithelial cells has been hindered by lack of appropriate tools and reagents, its importance in regulating *Dll4* expression has been unequivocally demonstrated. The *Dll4* and other Notch ligands secreted by TEPCs and TECs have been studied in the context of importance for T-cell development, but the importance of Notch signalling pathway in TEPCs and TECs has not been extensively studied. Mice mutant for Notch receptors exhibit medullary defects postnatally (Liu, O`Neill and Blackburn, unpublished). Interestingly, *Foxn1* is required to maintain *Notch1* expression in the hair follicle matrix (Cai et al. 2009). This study suggested that FOXN1 could bind to one or more of its consensus binding sites *in vitro* within the mouse *Notch1* promoter to regulate its activity. The up-regulation of *Notch1* expression at E12.5 suggests that similar to hair follicle matrix, the expression of *Notch1* in TEPCs could also be regulated by *Foxn1*. It would also be interesting to determine if *Notch1* is expressed homogeneously in all TEPCs or weather TEPCs represent a heterogeneous population with varied levels of *Notch*-signalling activities. Flow cytometry analysis using fluorescently labelled antibody against NOTCH1 or a fluorescent *Notch1* reporter allele could be used to answer this question.

One of the key limitations of this chapter is that only a small number of genes were analysed based on the list of candidate transcriptional regulators generated previously using bioinformatics approach. Thus, there is a possibility that the transcriptional regulator of *Foxn1* might be absent from this analysis if its expression does not correlate with that of *Foxn1* in the samples (E12.5 $MTS20^{+}$ and E15.5 $MTS20^{+}$ and $MTS20^{-}$ TECs) used for microarray based gene expression analysis. One way to overcome this would be to perform whole transcriptome analysis using microarray or RNA-seq during thymus development in WT mice, which can help identify candidate regulators with high confidence.

3.4 Summary

The focus of this chapter was to generate expression profiles, during development of both WT and *Foxn1*$^{-/-}$ thymus, of genes present in the list of candidate transcriptional regulators of *Foxn1*. The rationale for this work was that the expression of these genes has not previously been studied using quantitative gene expression approaches and a better understanding of this was required to identify whether these genes play a role in regulating *Foxn1* expression and to distinguish between transcriptional activators and targets of *Foxn1*. The results generated from these experiments have shed light on the dynamic nature of transcriptional regulation in developing thymus and have allowed identification of potential transcriptional interactions between key genes. The proposed transcriptional network diagram, shown in Figure 3.12, is the first to connect the genes known to be important for TEPC and TEC formation or function with those involved in the upstream network required for thymus primordium formation. The network diagram shows presence of feed-forward and feed-back mechanisms in the predicted genetic network. Of note is that the proposed transcriptional network is based only on gene expression analysis and further work is needed to determine whether these interactions are direct or indirect.

4. Identification of genome-wide transcriptional regulatory regions active in TEPCs by mapping histone modifications using ChIP-seq

4.1 Introduction

The gene expression analysis shown in Chapter-3 led to identification of potential genetic interactions in TEPCs. However, it did not allow for identification of transcriptional regulatory regions, which could in turn be used to test these interactions *in-vivo*. Genetic interactions have traditionally been characterised through generation of mutant mice either lacking or over-expressing a gene of interest. However, this does not allow differentiating between direct and indirect interactions between transcription factors and their potential target genes. Binding of transcription factors to DNA can be identified using Chromatin ImmunoPrecipitation (ChIP). This approach, which has been widely used in a number of cell types and model organisms, is limited by availability of ChIP-grade antibodies for the transcription factor under study and thus cannot yet be applied to all transcription factors without further genetic modification. As ChIP-grade antibodies for the transcription factors I wished to study further were not available, I chose as an alternative approach to identify active transcriptional regulatory regions on a genome-wide basis. This would allow subsequent testing of whether a particular transcription factor could bind to and regulate the function of a given gene or set of genes.

The temporally and spatially restricted transcription of genes is governed by binding of transcription factors to enhancers and promoters of the gene. The core promoter is defined as the minimal DNA sequence around the transcriptional start site (TSS) that accurately directs initiation of transcription by binding of core transcriptional machinery. On the other hand, enhancers can be located much further away, ranging from tens to hundreds or thousands of kilobases, from the gene and are thus more difficult to identify. Traditional approaches have focused on identification of enhancers for one gene at a time, however, the latest advancements in epigenetics

and next generation sequencing now allows for genome wide identification of a subset of putative enhancer regions. This combined with the ease of genome manipulations provided by the CRISPR-Cas9 system can be used to test a large number of identified regions for their function *in-vivo*. This approach has so far not been applied to TECs. Thus, it provided a unique opportunity for identification of molecular mechanisms governing thymus development. I therefore decided to apply this approach to identify genome wide promoters and enhancers for TEPCs.

To this end I used specific histone modifications to identify active promoters and enhancers in TEPCs, an approach that has been widely employed for this purpose in various cell types and model systems. It has been well established that active promoters and enhancers are flanked by histones containing H3K4me3-H3K27ac and H3K4me1-H3K27ac modifications respectively (Heintzman et al. 2007; Rada-Iglesias et al. 2011). Thus, mapping these modifications, most commonly achieved using ChIP-seq, on the genome of a particular cell type can be used to identify the set of candidate promoters and enhancers that are active in that cell-type. I, therefore, decided to perform ChIP-seq for these histone modifications on TEPCs isolated from the E12.5 thymus, since this population is the most homogenous population of fetal TEPCs currently identified, and is known to express high levels of *Foxn1*. The rationale behind this approach was that identification of these regulatory regions and the transcription factors that can bind to these (through bioinformatics approaches) could help filter the list of candidate transcriptional regulators of *Foxn1* for genes with highest probabilities of this function in TEPCs and also identify genes that could have been missed by the previous approach.

4.2 Optimisation of ChIP protocol for use with TEPCs isolated from E12.5 thymi

Shen and colleagues have previously generated histone modification profiles using 8-weeks old whole thymus (Shen et al. 2012). Given that the developing T-lymphocytes constitute the majority of cells (>95%) in an 8-weeks old thymus, this

data set is not of use for identification of TEC-specific histone modifications. Furthermore, histone modifications observed at TEC-specific genes in those data represent an average of modifications presents in various TEC sub-populations. A more recent study performed ChIP-seq for H3K4me3 and H3K27me3 on mature and immature mTEC subpopulations from 4-weeks old mice (Sansom et al. 2014). This study, however, did not analyse histone modifications in TEPCs and did not perform ChIP for histone modifications that can identify enhancers. To overcome these limitations, I decided to generate TEPC-specific histone modification profiles.

An important technical limitation to this approach was that, of necessity, it required working with samples obtained from a very low number of cells. We therefore obtained a ChIP protocoal optimised for low-cell (~100,000 cells) number ChIP-seq from the Wysocka lab (Stanford University). This number of TEPCs cells can be obtained from around 4 mice litters at E12.5. While enough TEPCs could be obtained for the actual ChIP-seq experiment, the availability of additional sample for protocol testing and optimisation was limited. Thus, the ChIP protocol was tested and optimised using ~100,000 E14/T embryonic stem cells (ES) cells. To ensure minimal difference in sample processing between TEPCs and the ES cells, I sorted and stored ~100,000 ES cells for subsequent ChIP, using equivalent protocols to those used for TEPCs. Figure 4.1 shows a flowchart representation of the optimised ChIP protocol. The first key step in ChIP, after fixation, is sonication of the fixed chromatin. The efficiency of downstream protocol is influenced by the size distribution of DNA fragments and the classic ChIP protocol includes a gel-based size selection prior to IP. Effective sonication was even more important when starting with ~100,000 cells as it was advisable to avoid sample loss resulting from gel purification of the sonicated chromatin. In these experiments, the sonication conditions resulted in DNA fragments between 100bp and 300bp for both E14/T and TEPCs (Figure 4.2), which is optimal for subsequent IP and sequencing library generation.

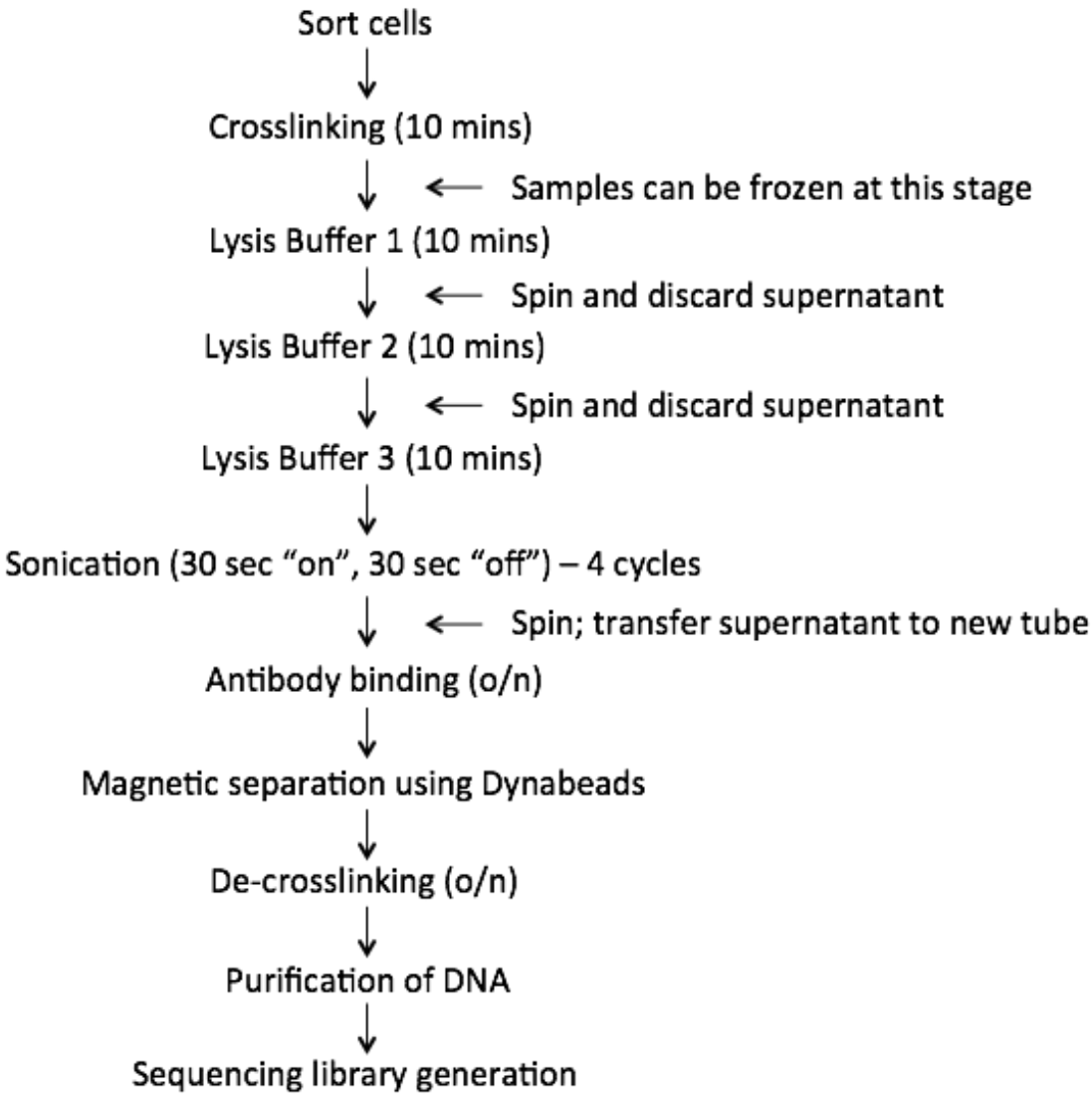


Figure 4.1: Schematic of ChIP-seq protocol
A schematic of the protocol used for chromatin immunoprecipitation of thymic epithelial progenitor cells. ~100,000 epithelial cells were used per sample. The entire protocol takes four days from sample collection to sequencing library generation.

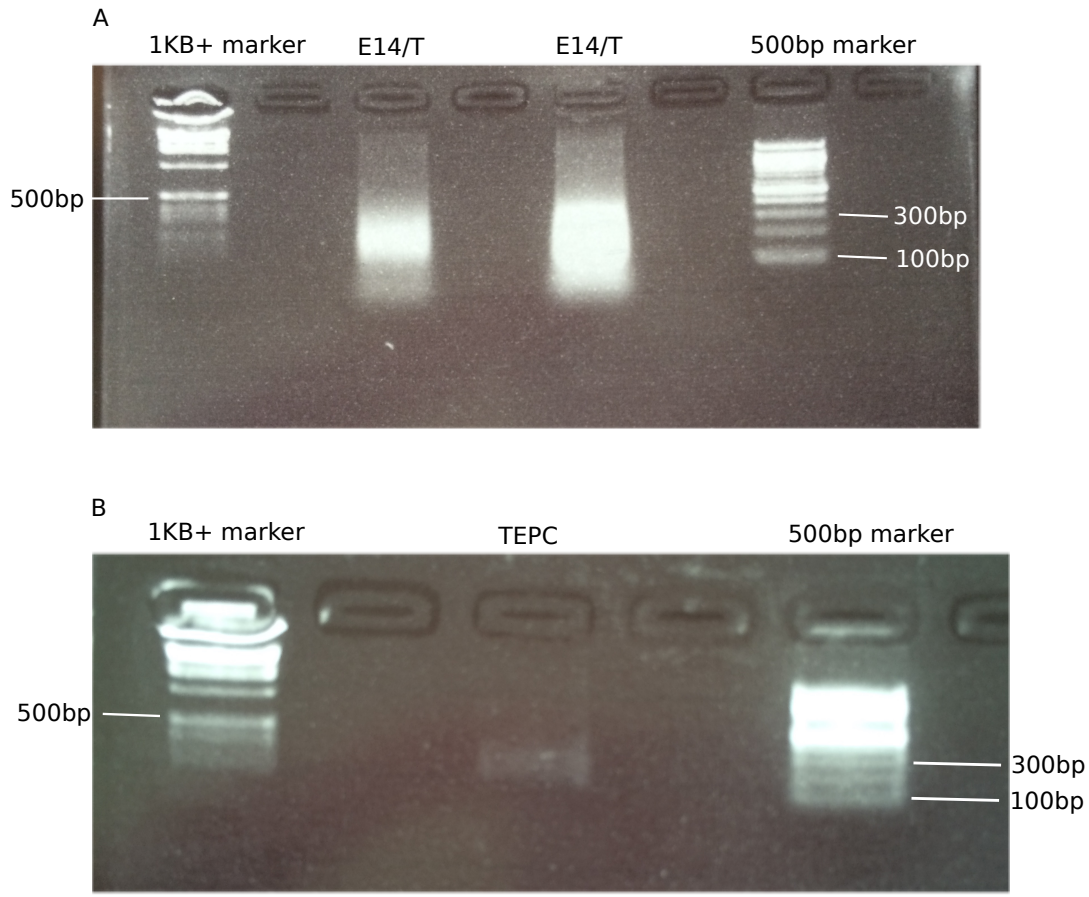


Figure 4.2: Sonication efficiency for ChIP samples.
~100,000 cells were sorted for ChIP. Four rounds of 10 sonication cycles (30sec "on", 30sec "off") were performed on formaldehyde crosslinked samples. Sonicated chromatin was purified from decrosslinked samples and analysed using agarose gel elcetrophoresis. Data shown are agarose gel images demonstrating the size of frgamented chromatin from (A) mouse ES cells: seven ES cells samples containing 100,000 cells each were pooled together after DNA purificaiton and (B) TEPCs: single sample containing 100,000 cells. A consistent band of the size 100-300bp was observed for these sonication conditions.

The immunoprecipitation (IP) was also optimised using E14/T ES cells. A detailed protocol for ChIP is described in Chapter 2. The success of IP for histone modifications was determined by real time-PCR of purified, de-crosslinked genomic DNA fragments, using primers against *Glyceraldehyde-3-phosphate dehydrogenase* (*Gapdh*) and *Actin, beta* (*b-actin*) promoters and *Nanog homeobox* (*Nanog*) and *Estrogen related receptor, beta* (*Esrrb*) enhancer regions known to contain the appropriate histone modifications (personal correspondence, Chambers group). Intergenic regions devoid of promoter or enhancer related histone modifications were chosen as negative controls (personal correspondence, Chambers group). As shown in Figure 4.3A and Figure 4.3B, a significantly higher enrichment was observed for positive control loci versus negative controls for both H3K4me3 and H3K27ac. Furthermore, the fold enrichment observed for positive control loci relative to IgG indicated that the efficiency of IP was suitable for generating samples for preparation of sequencing libraries.

Following establishment of the optimal IP conditions, I tested these conditions for TEPCs by performing ChIP-qPCR. The data shown in Figure 4.3C show that efficient IP was achieved for H3K4me3 in the TEPC samples. On the other hand, the fold enrichment observed for H3K27ac in TEPC samples was substantially lower than that observed in ES cells. Similar results were obtained for the first batch of ChIP-seq data (rep-1) generated from TEPCs (details below). This ChIP-seq rep-1 data showed excellent enrichment for H3K4me3 and H3K4me1 but suffered from low signal (enrichment) for H3K27ac. It has been suggested that acetylation of histone residues is more dynamic than methylation, and that endogenous deacetylase enzymes could affect the efficiency of H3K27ac IP. Thus, to determine whether the enrichment for H3K27ac in TEPC samples could to be improved through the use of histone deacetylase inhibitor, I decided to perform ChIP in the presence sodium butyrate (NaBu), a commonly used histone deacetylase inhibitor. NaBu was added at a concentration of 20mM to all solutions and buffers used for dissecting, sorting, formaldehyde fixation, and IP of TEPCs. Addition of NaBu had modest to substantial positive effect on the H3K27ac enrichment in TEPCs, as detected by ChIP-qPCR (Figure 4.3D). However, NaBu can also cause changes in transcription

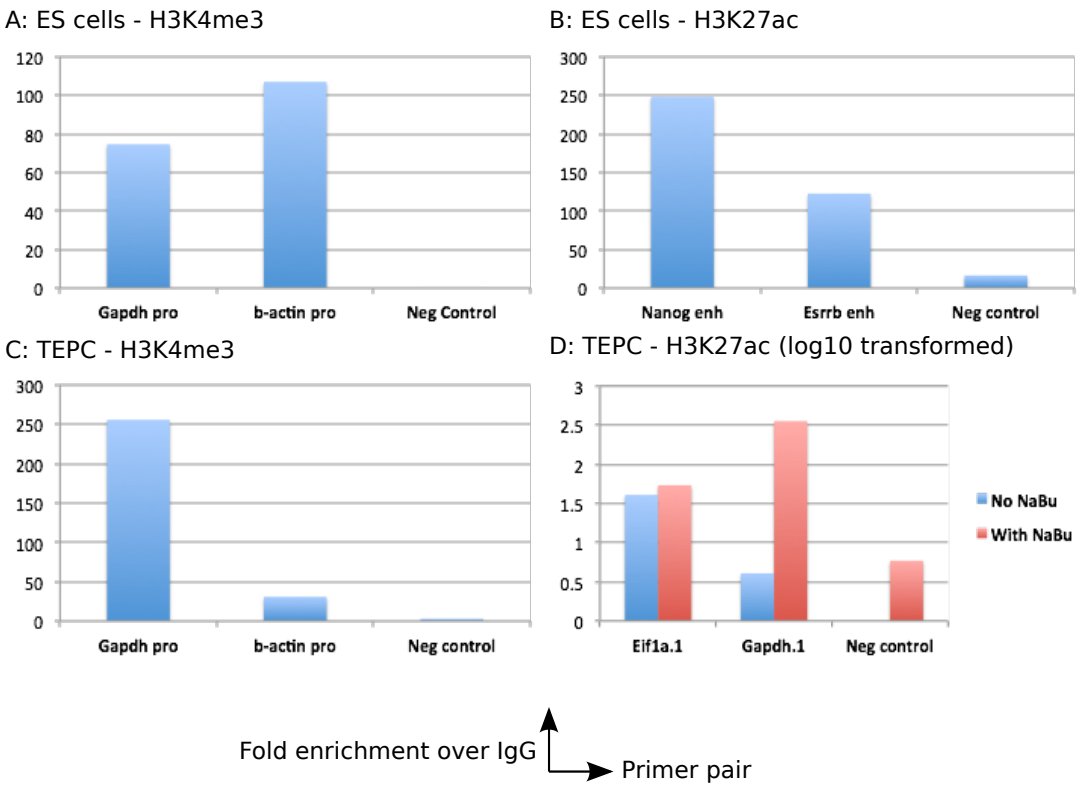


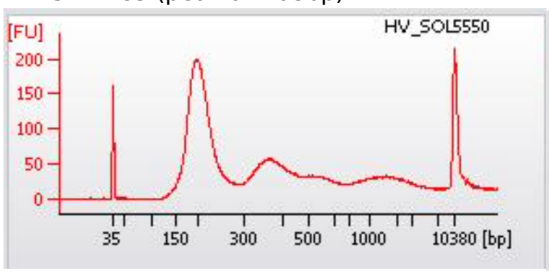
Figure 4.3: Enrichment of positive and negative control loci for H3K4me3 and H3K27ac following ChIP in ES cells and TEPCs.
ChIP was performed on ~100,000 ES cells and TEPCs and enrichment of immunoprecipitated DNA was determined using QRT-PCR. Data shown are fold enrichment over IgG control at positive and negative control loci for H3K4me3 in (A) ES cells (E14/T) and (C) TEPCs and for H3K27ac in (B) ES cells and (D) TEPCs. Fold enrichment for H3K27ac in TEPCs is shown both in presence and absence of histone deacetylase inhibitor, sodium butyrate (NaBu) (D). Data shown are representative of at least two biological replicates.

of genes when used for preventing histone deactylase activity in *in-vitro* cultures (Zhang & Wu 2013). Given that the TEPC samples were treated with NaBu during sample preparation stages for cell sorting (dissection, dissociation, staining, and cell sorting), when the cells are alive and could respond to presence of NaBu, an appropriate control was to perform panH3 ChIP on TEPCs treated in the same way. Thus, I decided to perform the second batch of ChIP-seq in presence of NaBu for all histone modifications. The rationale behind this approach was that treatment of the control sample with NaBu would normalise for any possible effect on histone occupancy. Thus, the second batch of ChIP samples, for H3K4me3, H3K27ac, and H3K4me1 histone modifications and pan-H3, were treated with 20mM NaBu during the entire ChIP protocol. The results from both batches of ChIP-seq are discussed below.

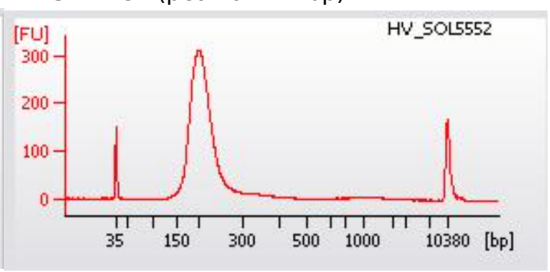
4.3 Generation of ChIP sequencing libraries from ~100,000 TEPCs

Following ChIP, sequencing libraries were generated using Diagenode's Microplex Library Preparation Kit which allows generation of sufficient quantity of high-quality sequencing library for good genome-wide coverage from very small amounts of DNA obtained by ChIP. The libraries were sent for sequencing to a commercial sequencing centre where the size distribution of the different libraries was determined using Agilent Bioanalyser. The size distribution traces for one set of ChIP-seq libraries are shown in Figure 4.4. As seen in this figure, the average size distribution for each library was between 200 and 230bp, which is optimal for sequencing on the Illumina HiSeq 2000 platform. The size distribution trace for the H3K4me3 sequencing library showed multiple peaks of higher fragment size (Figure 4.4A), which were most likely caused by slight over-amplification of the library. However, this was not expected to adversely influence the results and each set of libraries (one biological replicate for each of the histone modifications and panH3) was pooled and sequenced on one lane of HiSeq2000 sequencer. The libraries were sequenced to a depth of approximately 30 million reads per library for rep-1 and approximately 38 million reads per library for rep-2. Fastq files containing adapter-

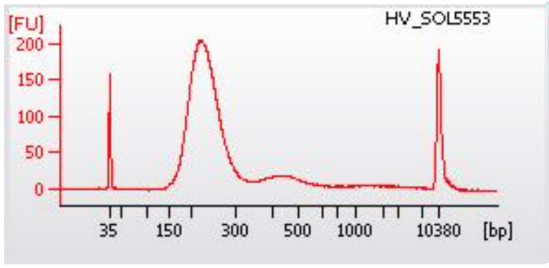
A: H3K4me3 (peak at 208bp)



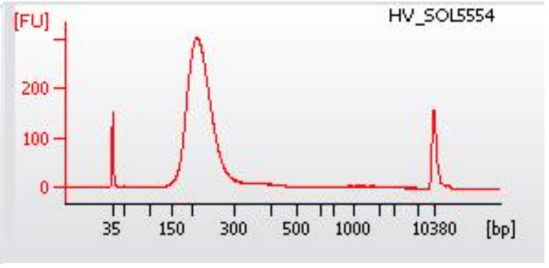
B: H3K4me1 (peak at 214bp)



C: panH3 (peak at 231bp)



D: H3K27ac (peak at 222bp)



E: pooled library (peak at 254bp)

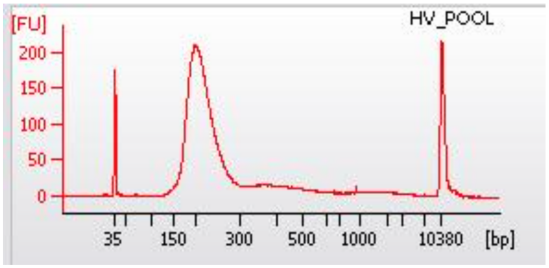


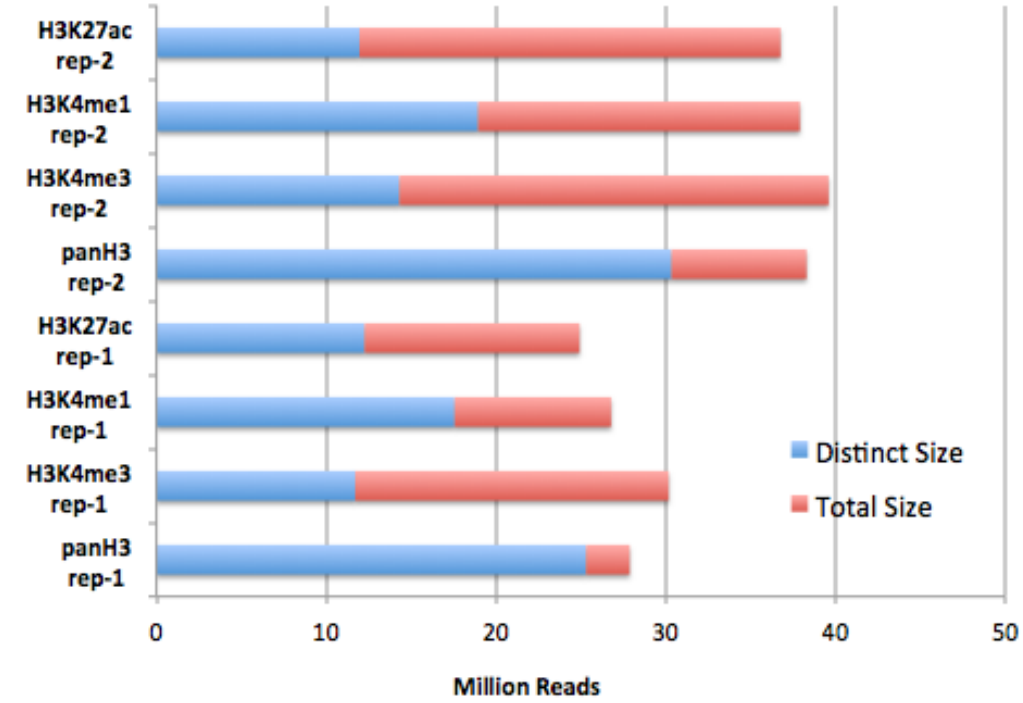
Figure 4.4: Bioanalyser traces of ChIP sequencing libraries
The size distribution of the ChIP sequencing libraries was determined using Agilent Bioanalyser. Data shown are size distribution traces of sequencing libraries for rep-1. (A) H3K4me3, (B) H3K4me1, (C) panH3, and (D) H3K27ac. (E) shows size distribution of the pooled library used for sequencing. Also shown are DNA fragment lenghts at which peaks are observed in these Bioanalyser traces (i.e. median frgament size for each sequencing library).

trimmed sequenced reads were obtained from the sequencing company for downstream bioinformatics analysis.

4.3.1 Quality control analysis of the sequenced reads

The data generated from sequencing of the ChIP-seq libraries was uploaded to GeneProf, an online platform for analysis of high-throughput sequencing data (Link: http://www.geneprof.org/GeneProf/show?id=gpXP_002962) (Halbritter et al. 2011). Sequenced reads quality control and alignment to mouse reference genome NCBIm37/mm9 were carried out using "Quality Control & Bowtie Alignemt" module using default parameters (Langmead et al. 2009; Halbritter et al. 2011). Figure 4.5 shows both the total and distinct sizes for each library. The total library size represents the number of sequenced reads for a library whereas the distinct size represents the number of distinct reads (differing by at least 1bp) sequenced. A comparable number of total reads was obtained between libraries within each batch (except panH3 rep-2, details below), indicating appropriate pooling of the libraries prior to sequencing (Figure 4.5). An important measure of the efficiency of IP and subsequent library generation is the presence of unique reads in the sequenced sample. Duplicate reads usually arise from PCR amplification bias or library generation from insufficient amount of starting material and can affect the reliability of peak identification (increased false positive) (Chen et al. 2012). The panH3 rep-1 library generated a very high proportion of distinct reads (~90%), whereas the panH3 rep-2 library generated a lower proportion of distinct reads (~80%). This was because the panH3 rep-2 library was sequenced at a higher sequencing depth (39M reads v/s 28M reads) than the panH3 rep-1 library, and increasing the sequencing depth increases the proportion of duplicate reads (Chen et al. 2012). The proportion of distinct reads for the histone modification ChIP-seq libraries ranged from 30-65%, possibly reflecting the differences in efficiency of IP and the abundance of these histone modifications within TEPCs, which would influence the amount of ChIP-generated DNA available as starting material for library generation in each sample.

A



B

Dataset	Distinct Size	Total Size	Ratio
panH3 rep-1	25,310,097	27,872,668	90.806
H3K4me3 rep-1	11,706,898	30,167,194	38.807
H3K4me1 rep-1	17,560,795	26,779,889	65.575
H3K27ac rep-1	12,258,407	24,922,032	49.187
panH3 rep-2	30,337,269	38,292,001	79.226
H3K4me3 rep-2	14,308,983	39,592,909	36.14
H3K4me1 rep-2	18,942,454	37,902,230	49.977
H3K27ac rep-2	11,954,612	36,763,629	32.517

Figure 4.5: Number of total and distinct sequenced reads for ChIP-seq samples.
Data shown are the number of total and distinct sequenced reads obtained for each ChIP-seq sample as (A) bar graph and (B) table. The number of total reads (total library size) represents the number of reads sequenced from a ChIP-seq library, whereas the number of distinct reads (distinct library size) represents the number of distinct (differing by at least 1bp) reads. Also shown is the ratio of distinct size to total size.

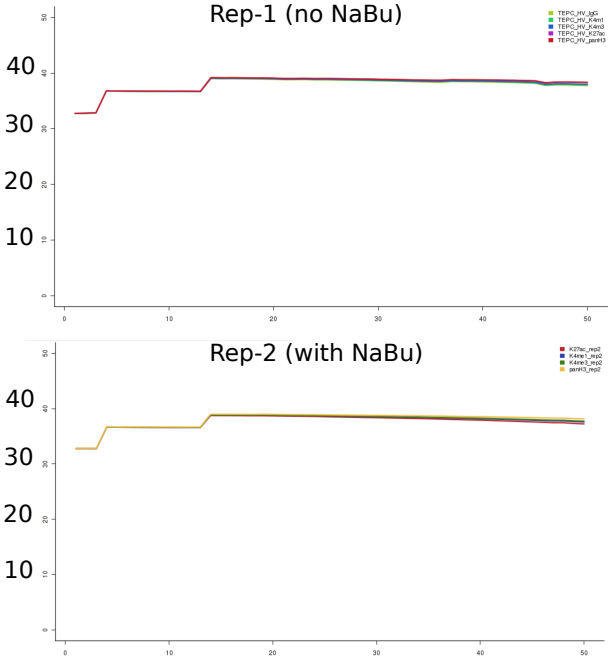
A measure of quality of sequenced reads is the Phred-like score generated during sequencing. The Average Quality Score Per (sequencing) Cycle data (average quality score along the length of sequenced reads; Figure 4.6A) and number of reads per quality scores (distribution of average quality score for each read; Figure 4.6B) indicated that high quality sequenced reads had been generated for both sets of sequencing libraries.

An important factor influencing the identification of peaks from a ChIP-seq sequencing library is its complexity. ENCODE defines library complexity as the fraction of DNA fragments that are nonredundant and suggests the use of complexity metric (nonredundant fraction, NRF), calculated as the ratio between the number of positions in the genome that uniquely mappable reads map to and the total number of uniquely mappable reads, to determine library complexity (Landt et al. 2012). A low complexity library is likely to yield little biological information due to increased proportion of redundant reads, which would be ignored during peak calling. The library complexity depends on a number of factors, including quality of antibody, number of cells used for IP, fixation and IP conditions, etc. ENCODE recommends obtaining sequencing libraries with NRF $\geq$ 0.8 for 10 million uniquely mapped reads, for point source libraries, such as the histone modifications studied here. To determine library complexities, I decided to perform a retrospective analysis on sequenced libraries by randomly sampling 10 million reads from each library (for consistency with ENCODE guidelines) and then aligning these subsample of sequenced reads using the parameter used for original libraries.

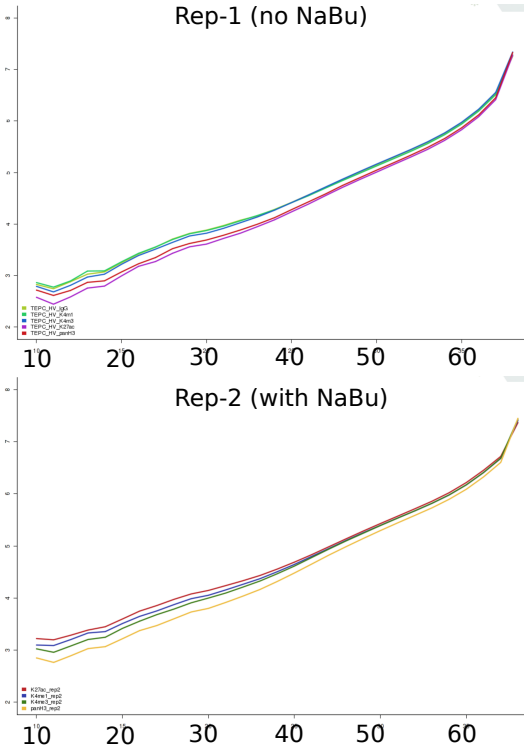
Figure 4.6: Quality control of sequenced reads and libraries

See following page
(A) A plot of quality score per sequencing cycle (i.e. per base pair position) averaged for all the reads within a sequenced library. (B) A plot showing the distribution of average quality scores for sequenced reads within a library. The quality scores shown represent Phred-like scores, which measures the probability that a base is called incorrectly. Higher scores represent a smaller probability of error. A quality score of 10 (indicating a 90% base call accuracy) is generally used as a cut-off. (C) The non-redundant fraction (NRF), calculated as per ENCODE guidelines, indicating sequenced library complexities. Graphs in (A) and (B) produced in GeneProf.

A



B



C

Sample	NRF
panH3 rep-1	0.98
H3K4me3 rep-1	0.62
H3K4me1 rep-1	0.83
H3K27ac rep-1	0.70
panH3 rep-2	0.95
H3K4me3 rep-2	0.61
H3K4me1 rep-2	0.75
H3K27ac rep-2	0.58

Since GeneProf currently does not allow creation of random samples with more than 250,000 reads, random sampling was carried out in R using the Bioconductor package ShortRead (Morgan et al. 2009). Three independent random samples of 10 million reads each were generated from each sequenced library. These were subsequently aligned and an average of their alignment statistics, i.e. number of distinct regions identified, was taken to determine NRF. As shown in Figure 4.6C, both the panH3 libraries have a very high NRF, indicating highly complex libraries. The NRF for the remaining libraries, except H3K4me1 rep-1, was lower than the recommended 0.8, suggesting that these libraries were less complex than that recommended by ENCODE. The lower complexities of these libraries are most likely the result of low-cell numbers used for ChIP, a limitation that cannot be easily overcome in this case. Interestingly, the NRF for H3K27ac sequencing libraries was lower when NaBu was used to prevent histone deacetylase activity, however this did not affect identification of peaks from this sample as described below.

4.4 Identification of peaks from ChIP-seq data

Following alignment of sequenced reads, uniquely aligned reads were used as input to the peak-calling algorithm MACS v1.4 (Zhang et al. 2008; Feng et al. 2012) for identification of peaks using default Geneprof parameters (described in Chapter 2) (Figure 4.7A). Despite the apparently low complexity of the H3K4me3 sequencing libraries, 23,348 and 12,613 peaks were identified for H3K4me3 rep-1 and H3K4me3 rep-2, respectively. Thus, the lower complexity of this sequencing library did not appear to have significant adverse effect on peak identification. Similarly, 33,842 and 21,490 peaks were identified for H3K4me1 from rep-1 and rep-2, respectively. On the other hand, the lower complexity H3K27ac sequenced libraries identified 168 peaks for rep-1 and 44,618 peaks for rep-2. This discrepancy in the number of peaks identified between the two replicates is discussed in detail below.

4.4.1 NaBu improves the ratio of reads count to genomic region for H3K27ac

As discussed above, the use of NaBu significantly increased the number of peaks identified for H3K27ac with only 168 peaks identified for the samples prepared without NaBu, as opposed to 44,618 peaks identified from the samples prepared in the presence of NaBu (using the appropriate panH3 control samples). Identification of peaks depends, to a large extent, on the number of reads aligned to a particular genomic location. A visual inspection of differences in number of reads aligned at any genomic location can be carried out using a genome browser. As an example, Figure 4.8A shows the alignment of H3K27ac sequenced reads at the *Pax1* locus in the presence and absence of NaBu. The top plot (dark blue track) shows the distribution of reads for the sample prepared without NaBu, while the bottom plot (green track) shows the distribution for the sample prepared with NaBu. As evident from comparing the two tracks, the distribution of the sequenced reads for both samples is highly similar, however the two tracks differ in the number of reads aligned at any genomic location, as evident from the scale on Y-axis. A similar pattern was observed for other genomic locations inspected visually by using the Genome Browser in GeneProf, two more examples are shown in Figure 4.8B and Figure 4.8C.This suggests that the increase in the number of peaks identified in the presence of NaBu is due to improved signal-to-noise ratio as a result of preventing histone deacetylation rather than changes in H3K27ac distribution. Another possible explanation for the differences observed between rep-1 and rep-2 would be technical variations associated with IP, which can affect subsequent library generation and peak detection. However, several factors support the conclusion that the observed differences are due to the positive effect of NaBu on maintaining histone acetylation. Firstly, results shown in Figure 4.3B suggests that the H3K27ac antibody and the ChIP protocol employed in this study are suitable for efficient IP. Secondly, the increased signal/noise ratio and number of peaks identified in presence of NaBu is consistent with the increased fold-enrichment observed for H3K27ac at positive control loci in ChIP-qPCR, as shown in Figure 4.3D. Therefore, I conclude that further analysis of sample prepared in NaBu was appropriate.

ChIP-seq library	No. of peaks identified	No. of associated genes
H3K4me3 rep-1	23,348	12,828
H3K4me1 rep-1	33,842	14,587
H3K27ac rep-1	168	285
H3K4me3 rep-2	12,613	8,399
H3K4me1 rep-2	21,490	10,905
H3K27ac rep-2	44,618	17,784

Figure 4.7: ChIP-seq peaks and comparison of panH3 samples

The number of peaks identified from each ChIP-seq library using MACS. Also shown are the number of genes associated with these peaks.

4.5 Identification of promoters active in TEPC by analysis of H3K4me3 and H3K27ac ChIP-seq data

4.5.1 Comparison of H3K4me3 rep-1 and rep-2

H3K4me3 ChIP data can be used to identify gene promoters (Heintzman et al. 2007; Rada-Iglesias et al. 2011). To this end, I first compared the data from the two H3K4me3 ChIP-seq samples described above. Peak calling was performed using MACS v1.4 as described in Chapter 2 (Zhang et al. 2008; Feng et al. 2012). The number of peaks identified in the H3K4me3 rep-2 (12,613; using panH3 rep-2 as control) dataset was lower than in H3K4me3 rep-1 (23,348; using panH3 rep-1 as control) (Figure 4.7A). Further analysis showed that this reflected a lower number of reads aligned at any genomic location in rep-2 than in rep-1 (values for Y-axis in examples shown in Figure 4.9A and Figure 4.9B). To determine the similarities between H3K4me3 rep-1 and H3K4me3 rep-2, I compared the genes associated with peaks from both these samples. To this end, H3K4me3 peaks were assigned to a gene if they were within 1000bp of the gene's transcriptional start site or if the peaks overlapped with either exons or introns of the gene. This resulted in 12,828 and 8,399 genes being associated with peaks in H3K4me3 rep-1 and rep-2 respectively. Approximately 92% of the genes associated with peaks in H3K4me3 rep-2 were also associated with peaks in H3K4me3 rep-1, indicating that peaks from both these samples are located within promoter regions of similar sets of genes (Figure 4.10A). Furthermore, analysis of shared peaks (at least 1bp overlap) between H3K4me3 rep-1 and rep-2 showed that the number of tags at these shared peaks was higher in rep-1 than in rep-2 (Figure 4.10B), suggesting lower signal/noise ratio in rep-2. These differences between H3K4me3 rep-1 and rep-2 could either be an effect of the use of NaBu during dissociation and subsequent ChIP protocol or represents biological and/or technical variations between samples. However, more biological replicates are required to determine unambiguously between these alternative possibilities. Given that H3K4me3 rep-1 showed higher number of tags than H3K4me3 rep-2 at shared peaks, I decided to use H3K4me3 rep-1 for further analysis.

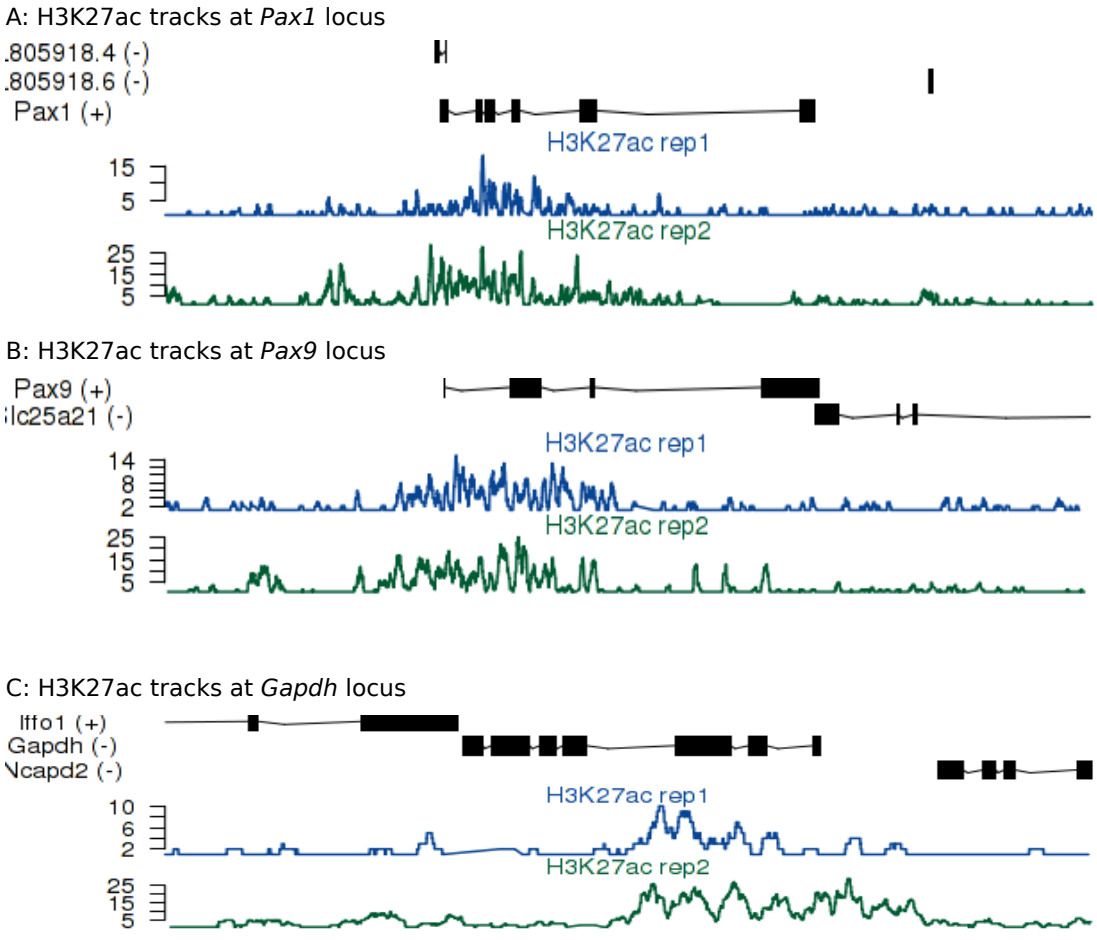


Figure 4.8: Comparison of H3K27ac rep-1 and rep-2
(A), (B), and (C) Comparison of genome browser traces for H3K27ac rep-1 (no NaBu) and rep-2 (with NaBu) at (A) *Pax1*, (B) *Pax9*, and (C) *Gapdh* genomic loci. Note the difference in the Y-axis values between the two samples, indicating higher number of reads aligned in rep-2 compared to rep-1 at each genomic loci. Genome browser tracks produced using GeneProf.

4.5.2 Identification of active promoters using H3K4me3 rep-1 and H3K27ac rep-2

Given that H3K4me3 rep-1 and H3K27ac rep-2 showed a good signal/noise ratio, I used these samples to identify active promoters. As expected, the majority (81%) of the H3K4me3 peaks were located within either "narrow" or "wide" promoter of genes (Figure 4.11A), as defined by GeneProf. On the other hand, since H3K27ac is not restricted to promoters, the peaks for this histone modification were widely distributed across the genome (Figure 4.11A).

To identify peaks associated with gene promoters, overlapping H3K4me3 and H3K27ac peaks were assigned to genes if they were within 1kb of the transcriptional start site (TSS), including all alternate TSSs. Promoters were classified as transcriptionally "active" if they contained overlapping H3K4me3 and H3K27ac histone modifications (Heintzman et al. 2007; Rada-Iglesias et al. 2011). In this analysis, peaks were considered to be overlapping if they shared at least 1bp. Using these criteria, 11,258 genes were identified as having active promoters in E12.5 TEPCs. Figure 4.11B and Figure 4.11C show two examples of genes with H3K4me3 and H3K27ac histone modifications. Figure 4.11D shows an example of a gene with H3K4me3 modification but no significant enrichment for H3K27ac modification, suggesting that the promoter may not be actively transcribing. These analyses provided a genome-wide indication of promoters that were likely to be active in TEPCs. Further analysis of the promoters of a selection of genes known to be important for TEPCs is described below.

4.6 Identification of putative active enhancers using H3K4me1 and H3K27ac

Similar to H3K4me3, more peaks were identified for H3K4me1 rep-1 than for H3K4me1 rep-2. Thus, H3K4me1 rep-1 was used for identification of putative active enhancers. Figure 4.12A shows the distribution of H3K4me1 rep-1 peaks relative to gene bodies. As expected, the location of H3K4me1 peaks showed a wide

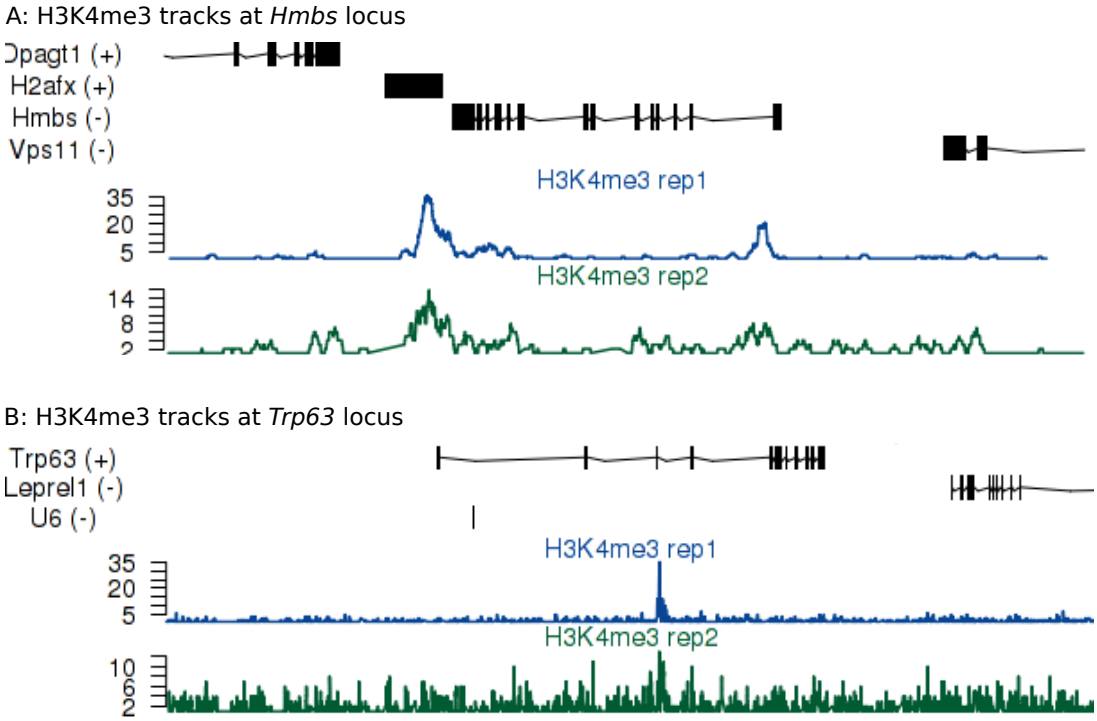


Figure 4.9: Comparison of genome browser tracks for H3K4me3 rep-1 and rep-2
Genome browser tracks comparing read densities for H3K4me3 rep-1 and rep-2 at (A) *Hmbs* and (B) *Trp63* genomic loci. Note the differences in Y-axis values.

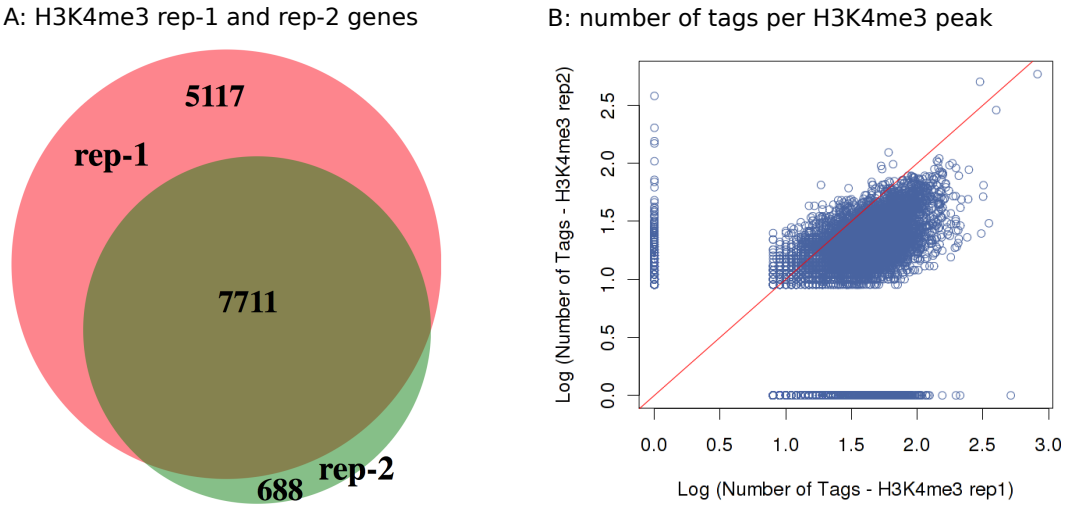


Figure 4.10: Comparison of H3K4me3 rep-1 and rep-2 peaks
(A) Overlap of genes associated with H3K4me3 peaks in rep-1 and rep-2. (B) Comparison of number of tags associated with peaks shared between H3K4me3 rep-1 and rep-2.

distribution, with only 22.18% located within "narrow" and "wide" promoters, 34.9% located within gene exons or introns, and the remaining peaks distributed such that they do not overlap any gene.

To identify putative active enhancers, the following workflow was applied: regions with overlapping H3K4me1 and H3K27ac peaks were identified and were required not to overlap with H3K4me3 peaks. Overlapping H3K4me1 and H3K27ac peaks passing this filter were then assigned to genes if they were within a 100kb window centred on transcriptional start sites. Peaks were assigned to only the nearest genes to avoid assigning a peak to multiple genes. Using the above criteria, 7,755 regions containing overlapping H3K4me1 and H3K27ac peaks were identified as active enhancers. These peaks mapped to 4,192 genes using the assignment criteria mentioned above. Figure 4.12B and Figure 4.12C show two examples of enhancers identified by this approach and their closest genes.

4.6.1 Identified putative enhancers are associated with genes important for thymus development

To investigate whether the putative enhancers identified in the above analysis were associated with genes with particular classes of functional annotations, I used the genomic regions corresponding to the identified peaks for ontology analysis with Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al. 2010). Genomic regions were assigned to genes using the default gene regulatory domain definition used by GREAT (basal plus extension). In this assignment approach, a basal region of 5kb upstream + 1kb downstream of TSS is assigned to each gene. The basal regions for a gene is then extended in both directions to the nearest gene's basal domain but no more than 1000kb in any direction (McLean et al. 2010). The genomic regions falling within this extended regulatory domain of a gene are assigned to that gene. This analysis revealed that the majority of regions (~85%) were associated with two genes (Figure 4.13A). The assignment of regions to only the closest gene, as described in the workflow above would result in loss of information when using the resulting associated genes list for ontology analysis.

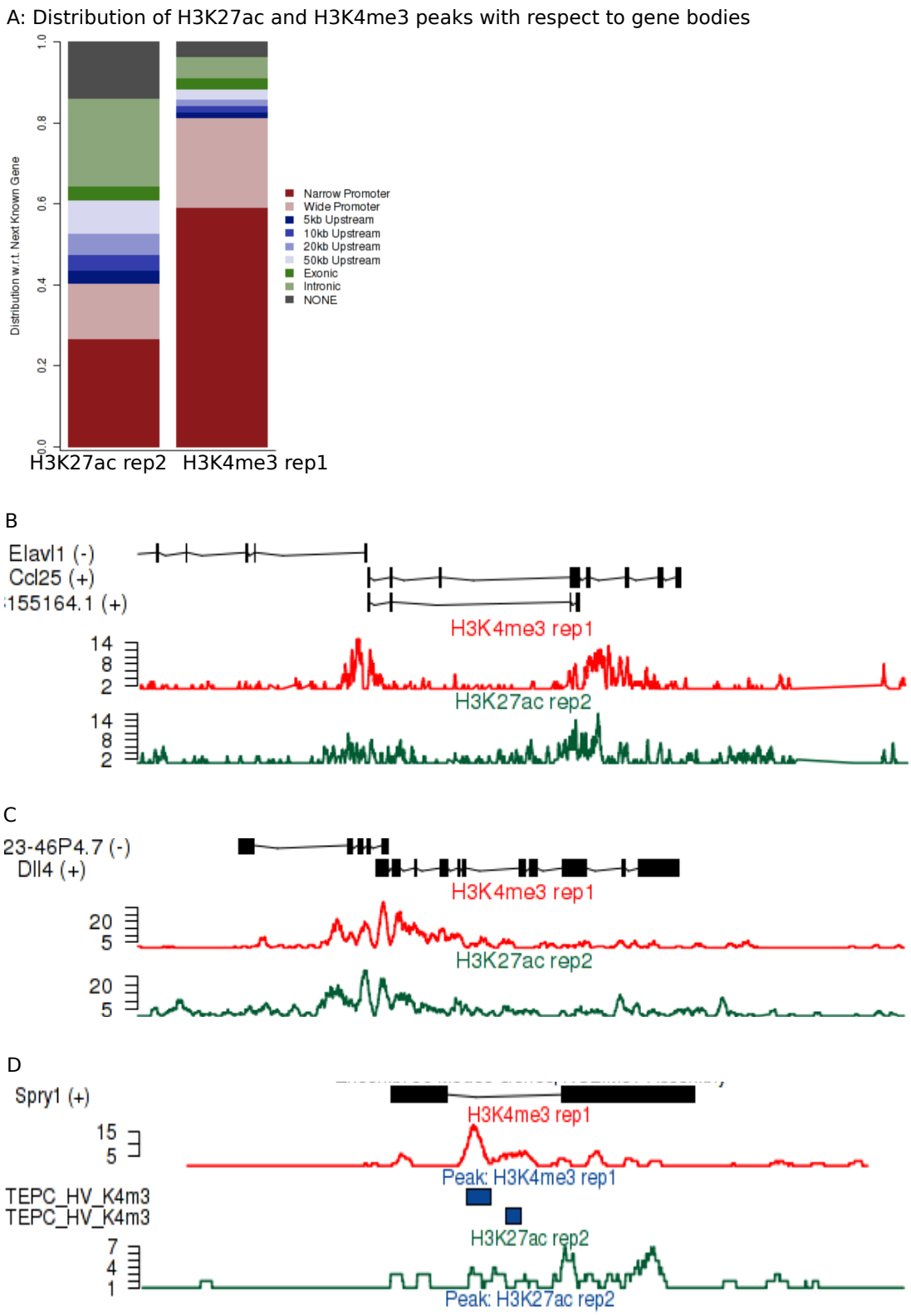


Figure 4.11: Identification of active promoters using H3K4me3 and H3K27ac
(A) Distribution of H3K27ac rep-2 and H3K4me3 rep-1 peaks with respect to gene bodies. (B), (C), and (D) H3K27ac rep-2 and H3K4me3 rep-1 tracks at *Ccl25* (B), *Dll4* (C), and *Spry1* (D) genomic loci showing enrichment of these histone modifications at gene promoters.

Therefore, where a putative enhancer was associated with two genes, both were included in the GREAT analysis discussed below. Furthermore, analysis of the distribution of the regions relative to the TSS of associated genes (Figure 4.13B) showed that more than half the regions were at least 50kb away from TSS, thus supporting the use of default association parameters of GREAT.

The genes associated with identified enhancer regions showed the highest enrichment for the Mouse Phenotype term: "thymus hypoplasia" (based on MGI phenotype ontology) (Figure 4.13C). This suggested that the ChIP-seq data generated here had successfully identified enhancers for genes important for thymus and T-cell development, as a defect in the later process can also lead to a hypoplastic thymus (Table 4.1). It is important to note that no peaks for either of the histone modifications under study were assigned to the *CD45* gene, which is expressed in T-cells but not in TECs. This suggests that the samples used for ChIP-seq did not contain any thymocytes and the enrichment observed here represents enrichment in TEPCs. Interestingly, among the enriched GO terms for Biological Process were terms such as "hair follicle development" and "hair cycle" (Figure 4.13D and Table 4.1) suggesting similarities between the development of TEPCs and hair follicle. Since *Foxn1* is expressed in both TEC and in hair follicle epithelium, this might reflect common genetic targets of FOXN1 in these cell types. Other enriched GO terms for Biological Process included "somatic stem cell maintenance" and "regulation of stem cell differentiation" (Figure 4.13D and Table 4.1), supporting the progenitor state of TEPCs and the initiation of their differentiation and potentially identifying genes involved in these processes; plus several terms associated with cell junctions, possibly reflecting the importance of genes associated with cell-cell junctions in the dramatic changes in structure associated with thymus organogenesis at E12.5. However, the association of genes to GO terms are based on different evidences and thus the importance and roles of genes associated with the above GO terms in thymus development need to be tested experimentally.

A: Distribution of H3K27ac rep-2 and H3K4me1 rep-1 peaks with respect to gene bodies

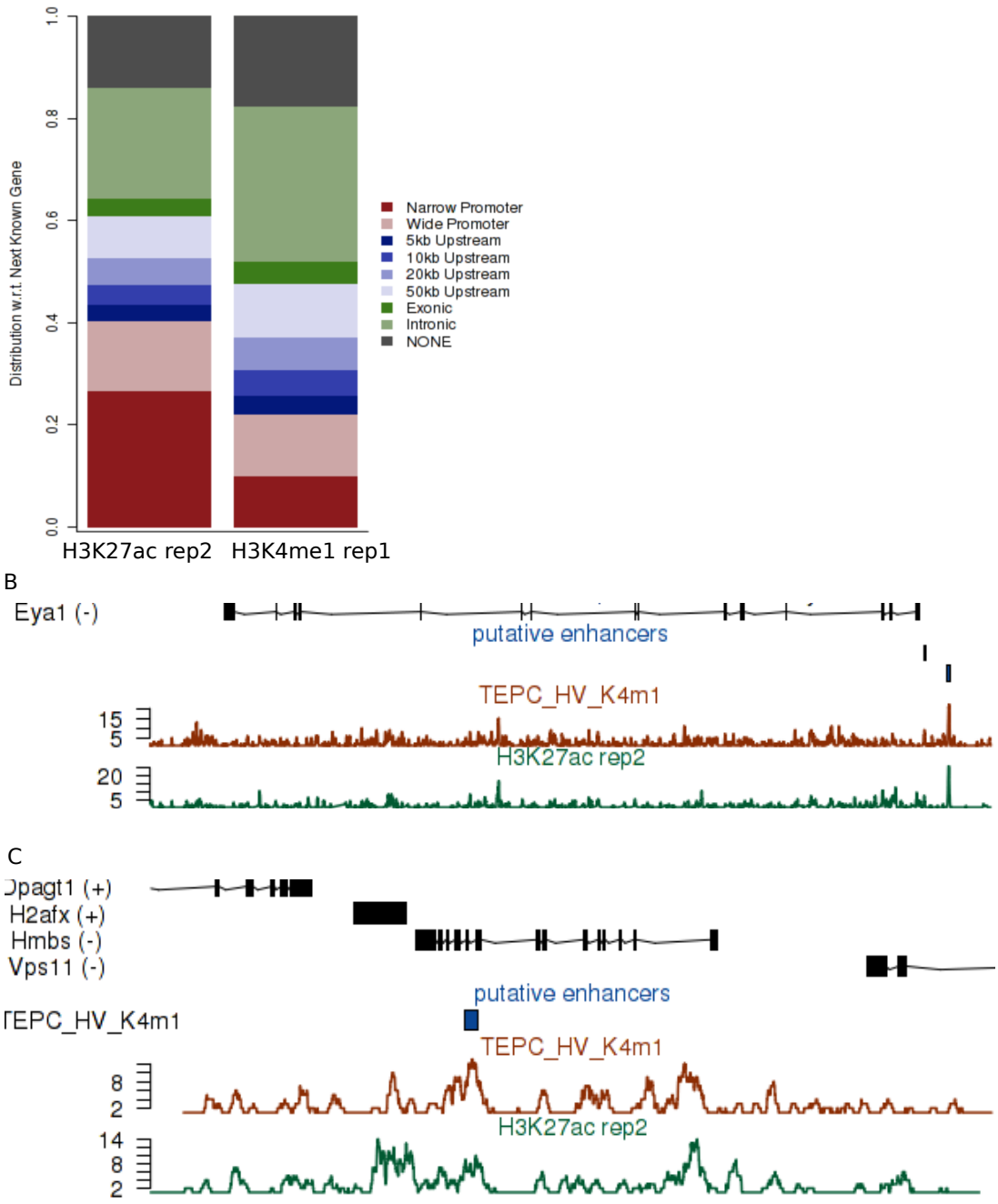


Figure 4.12: Identification of active enhancers using H3K4me1 and H3K27ac
(A) Distribution of H3K27ac rep-2 and H3K4me1 rep-1 peaks with respect to gene bodies. (B) and (C) H3K27ac rep-2 and H3K4me3 rep-1 tracks at (B) *Eya1* and (C) *HMBS* genomic loci showing enrichment of these histone modifications at putative enhancers..

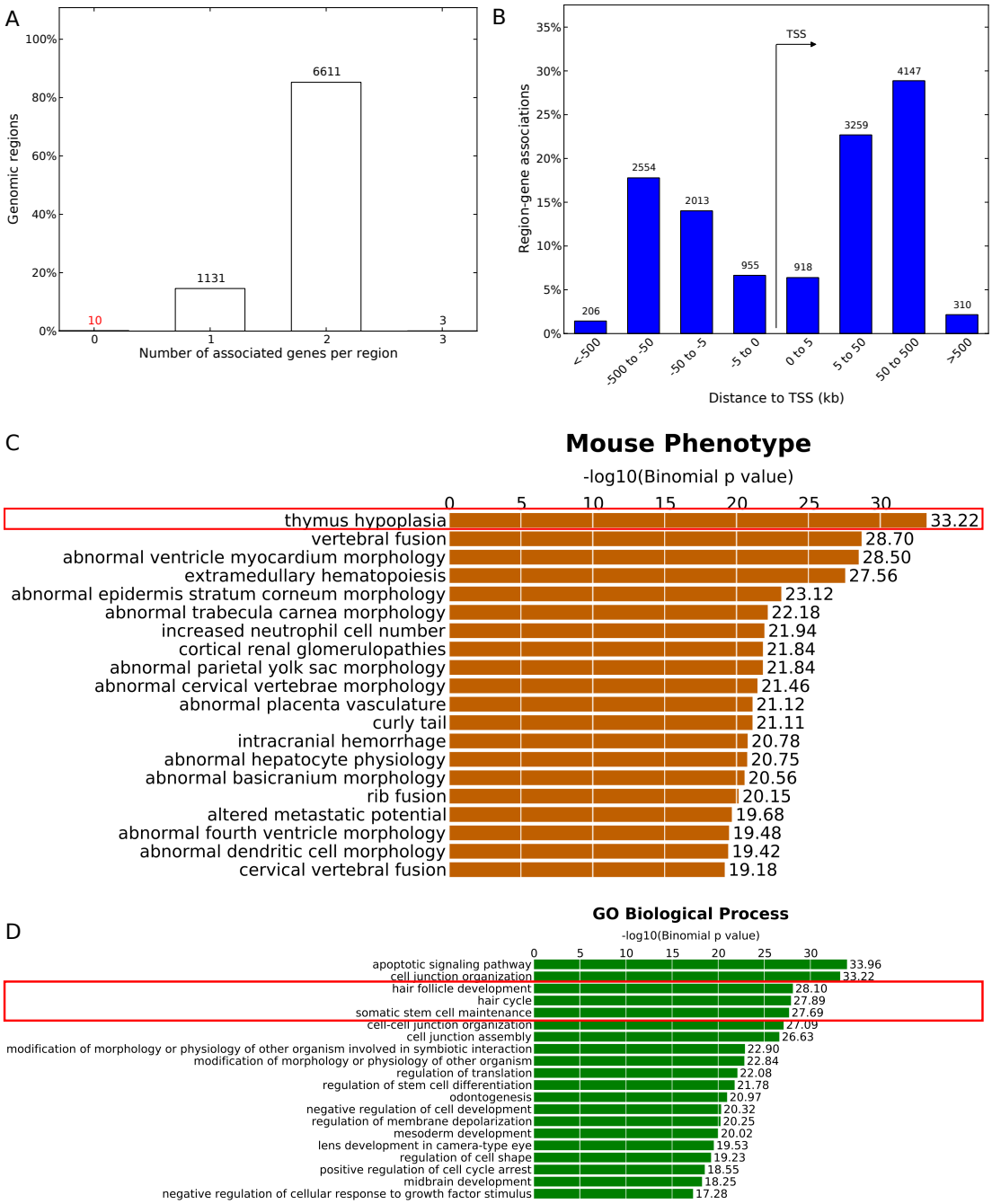


Figure 4.13: Analysis of putative enhancer regions using GREAT(1): mouse phenotype and biological process

The putative enhancer identified using H3K4me1 and H3K27ac data were analysed using the bioinformatics tool GREAT. Enhancers were assigned to genes using the default GREAT parameters that assigns a genomic region (enhancer regions in this case) to a gene if its wihtin a certain distance of the gene (distances varies for each gene). (A) The number of genes associated with each region. Most of the enhancer regions were assigned to two genes. (B) Distribution of distances of enhancer regions to the transcription start sites (TSS) of the assigned genes. The enhancers regions are spread over large distances from TSS. (C) Enrichment of genes that are assigned to enhancer regions for Mouse Phenotype GO terms. Highest enrichment observed was for genes associated with thymus hypoplasia phenotype. (D) Enrichment of Biological Process GO terms. Note the enrichment for terms associated with skin development.

Thymus hypoplasia (1)	Thymus hypoplasia (2)	Hair follicle development	Somatic stem cell maintenance	Stem cell differentiation
Abl1	Lbr	Acvr1b	Apc	Bambi
Adam10	Lck	Akt1	Bcl9	Bcl9l
Aff1	Mcl1	Apc	Bcl9l	Bmp7
Akt1	Mdm2	Barx2	Braf	Cnot2
Akt2	Msi2	Bcl2	Fzd7	Ctnnb1
Apc	Nfat5	Cdc42	Gata2	Dab2ip
Arid5b	Nfatc3	Celsr1	Hes1	Efna1
Bad	Nfkb1	Ctnnb1	Hmga2	Ell3
Bag3	Notch1	Ctsl	Kat6a	Epha3
Bcl11a	Nr3c1	Dicer1	Ldb2	Foxa2
Bcl11b	Pbx1	Edar	Lrp5	Gata6
Birc5	Pdgfra	Edaradd	Nog	Grem1
Cacna2d2	Pik3cd	Egfr	Sfrp1	Gsk3b
Cbx4	Pknox1	Fgfr2	Ski	Hes1
Ccnd3	Prkch	Foxe1	Sox2	Hmga2
Cd247	Prkdc	Foxn1	Sox4	Hpn
Cd3e	Prnp	Fst	Sox9	Jag1
Cdk6	Ptprc	Fzd3	Tcf7l1	Jup
Chd4	Rb1	Hdac1	Tcf7l2	Lef1
Chd7	Rorc	Hoxc13	Tfap2c	Notch1
Col10a1	Rpl22	Inhba	Vangl2	Pax8
Crip3	Runx1	Krt71	Wnt7a	Pdgfra
Cux1	Sgpl1	Lama5	Wnt9b	Pofut2
Dll4	Shc1	Ldb2	Zfp36l2	Ppp2ca
E2f2	Sit1	Lrp4	Zhx2	Prickle1
Ednra	Slc46a2	Mreg		Rgcc
Efnb2	Smad3	Ngfr		Sfrp1
Esr1	Sox4	Notch1		Sfrp2
Esr2	Src	Pdgfa		Smad2
Ets1	Ssbp2	Ppard		Smad3
Fbln1	Stk11	Psen1		Snai1
Fbxw7	Tcf12	Rela		Sox9
Fgf8	Tcf7	Runx1		Tbx5
Foxn1	Tgfb1	Runx3		Tcf7l1
Fyn	Trat1	Snai1		Tgfb1
Hdac1	Trp53bp1	Sox9		Tgfb2
Hes1	Upf2	Tfap2c		Tgfb3
Hnrnpl	Vav2	Tgfb2		Tgfbr3
Hoxa1	Vegfa	Tnfrsf19		Zfp36l2
Il10	Wipf1	Trp63		Zfp703
Il7	Xrcc4	Vangl2		
Jarid2	Xrcc5	Wnt10a		
Kat6a	Zap70	Zdhhc21		
Kit	Zbtb17			
	Zfp36			

Table 4.2: Genes contributing to enrichment of Mouse Phenotype and Biological process terms in results from GREAT analysis.

The genes contributing to the enrichment of the following terms in results from GREAT analysis: "thymus hypoplasia" in Mouse Phenotype; "hair follicle devleopment", "somatic stem cell maintenance", and "stem cell differentiation" in Biological Process. Shown here are genes that were assigned one or more enhancer regions and contributed to enrichment.

Analysis of genomic regions using GREAT also allowed identification of enriched signalling pathways using various pathway datasets. The use of one such datasets, MSigDB Pathway, indicated a strong enrichment for, among others, TGFβ signalling, BMP signalling, Notch signalling, p38 MAPK signalling, and NFκB signalling (Figure 4.14A), all of which have been implicated in thymus development or function. Terms associated with TGFβ and SMAD also showed strong enrichment in Molecular Functions ontology (Figure 4.14B). Together, this suggested that TGFβ signalling, through its downstream SMAD-dependant and p38 dependant functions, might play an important role in regulation of TEPCs. The enrichment of Notch signalling pathway is consistent with the upregulation of *Notch1* gene observed in Chapter 3. Similarly, the other enriched pathways too are likely to be important for thymus development. Of note is that the use of a different pathway dataset, PANTHER pathway, indicated enrichment only for the Notch signalling and p38 MAPK signalling pathways (Figure 4.14C). The genes with assigned enhancer regions and associated with each of these enriched terms are shown in Table 4.2. Finally, it is also important to note that while the GO and Pathway terms mentioned above are manually curated, the relevance of the genes associated with these terms and the function of the respective pathways in the cell-type under study needs to be experimentally validated.

4.7 Identification of *Foxn1* promoter and enhancers

Given the importance of *Foxn1* in thymus development and maintenance, I focused on identification of putative promoter and enhancers for this gene. Promoter and enhancers were identified from ChIP-seq datasets as described above. A detailed analysis of these regions is mentioned below:

4.7.1 *Foxn1* promoter

Figure 4.15A shows H3K4me3 rep1 and H3K27ac rep2 tracks at the *Foxn1* locus and the region designated as the active promoter for *Foxn1*. This analysis identified a second region within the first intron marked by both H3K4me3 and H3K27ac,

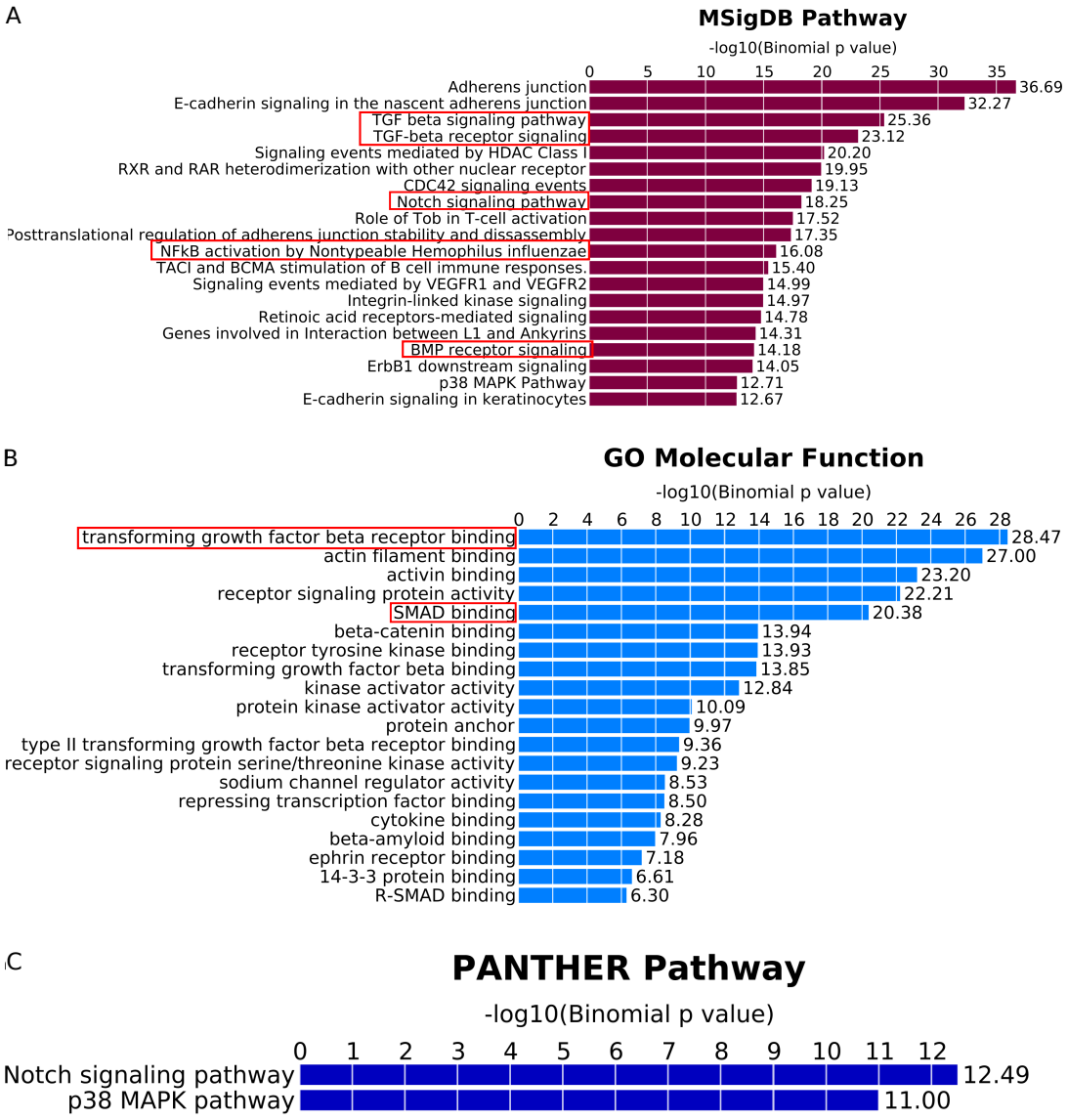


Figure 4.14: Analysis of putative enhancer regions using GREAT(2): signalling pathway and molecular function.

The putative enhancer identified using H3K4me1 and H3K27ac data were analysed using the bioinformatics tool GREAT. Enhancers were assigned to genes using the default GREAT parameters that assigns a genomic region (enhancer regions in this case) to a gene if its wihtin a certain distance of the gene (distances varies for each gene). (A) Enrichment of genes that are assigned to enhancer regions for MSigDB Pathways. MSigDB provides curated lists for genes involved in signalling pathways. Highlighted are enrichments for TGFbeta, Notch, NF-kB, and BMP signalling due to their importance in the thymus. (B) Enrichment of Molecular Function GO terms. Note the enrichment for terms associated TGFbeta signalling and SMAD binding, consistent with enrichment of this signalling pathway. (C) Alternative pathway enrichment analysis using PANTHER pathway.

TGFβ signalling (MSigDB)	BMP signalling (MSigDB)	Notch signalling (MSigDB)	p38 signalling (MSigDB)	NFkB activation by H. influenczae
Apc	Bambi	Aph1a	Akt1	Crebbp
Cdh1	Bmp4	Crebbp	Atf1	Ep300
Crebbp	Bmp7	Ctbp1	Cdc42	Map2k6
Ep300	Bmpr1b	Ctbp2	Creb3	Map3k14
Map3k7	Chrd	Dll4	Creb5	Map3k7
Skil	Ctdsp1	Dtx1	Dusp10	Mapk11
Smad2	Ctdsp2	Dtx3	Eef2k	Mapk14
Smad3	Ctdspl	Dtx4	Eif4e	Nfkb1
Smad7	Fst	Ep300	Hspb1	Nr3c1
Tab1	Grem1	Hdac1	Map2k6	Rela
Tgfb1	Gsk3b	Hes1	Map3k4	Smad3
Tgfb2	Map3k7	Jag1	Map3k4	Tgfbr1
Tgfb3	Nog	Jag2	Map3k5	Tgfbr2
Tgfbr1	Ppm1a	Lfng	Map3k7	Tlr2
Tgfbr2	Ppp1r15a	Maml1	Mapk11	
	Rgma	Maml3	Mapk13	
	Rgmb	Ncor2	Mapk14	
	Ski	Notch1	Mapkapk2	
	Smad1	Notch2	Mknk2	
	Smad5	Notch3	Nfkb1	
	Smad6	Numb	Srf	
	Smad7	Psen1	Tab1	
	Smurf1		Tab2	
	Tab1		Traf6	
	Tab2			

Table 4.3: Genes contributing to enrichment of signalling pathways in results from GREAT analysis.

The genes contributing to the enrichment of the following signalling pathways in results from GREAT analysis: TGFbeta, BMP, Notch, p38, and NF-kB signalling. Shown here are genes that were assigned one or more enhancer regions and contributed to enrichment.

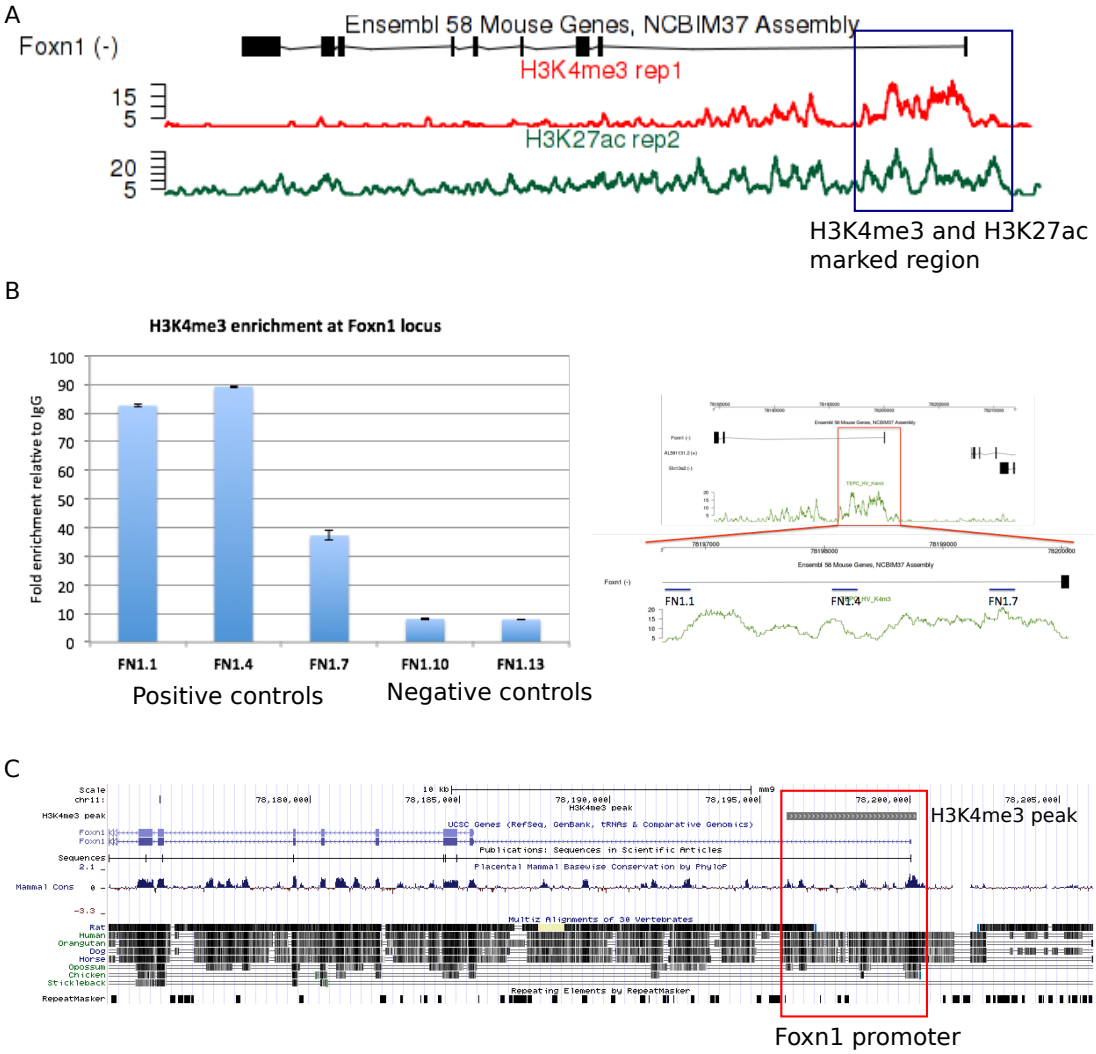


Figure 4.15: Identification of *Foxn1* promoter

(A) Genome browser tracks for H3K4me3 and H3K27ac showing the active *Foxn1* promoter. The enrichment of these histone modifications at a second region wiht intron-1 is likely to indicate limited transcription from alternative first exon of Foxn1. (B) ChIP-qPCR validation of H3K4me3 enrichment at active *Foxn1* promoter. Data shown indicate fold enrichment over IgG control. Also shown are genomic regions ampliefied by each positive control primer pairs. Negative control primer pairs amplified regions outside the identified promoter. Error bars indicate technical variability of qPCR. (C) UCSC browser image of the identified *Foxn1* promoter, represented by H3K4me3 peak. Also shown is track for mammalian conservation. The red box indicates the genomic region used as *Foxn1* promtoer for subsequent analysis, which was exentended to include conserved regions on either side of H3K4me3 peak.

indicating the presence of second promoter region. This promoter region most likely corresponds to exon-1b, an alternate first exon, of *Foxn1*, which has been reported in the literature and corroborated by our own studies, but is not curated in any genomic database. Transcripts utilising exon-1b are found in skin, where as thymic epithelial cells shown only minimal expression of this transcript with most of the *Foxn1* transcript starting at exon-1a. The presence of the second promoter region in this dataset suggests that transcripts starting at exon-1b might be made in TEPCs, albeit at much lower rate than those starting at exon-1a. The absence of mature transcripts starting at exon-1b in TECs further suggests that transcripts generated using the second promoter are likely to be prematurely terminated, leading to their degradation. The enrichment of H3K4me3 across the identified peak region was validated using ChIP-qPCR (Figure 4.15B). Based on these analyses I concluded that the H3K4me3 marked region as *Foxn1* promoter and further extended to include the highly conserved region to the 5` end (negative strand) of this region (Figure 4.15C). Thus, the coordinates of the genomic region defined as the *Foxn1* promoter in all subsequent analyses were as follows: (in mm9/GRCm37) Chr11: 78196637 – 78200589 or (in mm10/GRCm38) Chr11: 78383135 – 78387087.

I then tested the above-defined *Foxn1* promoter for TFBS using the Match programme, with parameters set to minimize the sum of false positive and false negative identification.

Table 4.3: Summary of transcription factor binding sites identified in *Foxn1* promoter

See following page
Foxn1 promoter was analysed for presence of transcription factor binding sites using Match algorithm and TRANSFAC professional database. This table lists the matrices, the associated transcription factors, and number of sites identified for each matrix.

Table 4.3 (1/2)

Matrix	Factor	Classification	Number of sites
V$EKLF_Q5_01	EKLF	ZFC2H2	1
V$SRF_Q5_02	SRF	MADS	1
V$NFY_Q3	NF-Y	HISTONE	1
V$RORALPHA_Q4	RORalpha	ZFC4-NR	1
V$NFAT1_Q4	NF-AT1	REL	1
V$SRY_Q6	SRY	HMG	1
V$REVERBALPHA_Q6	Rev-ErbAaLPHA	CHCH	1
V$RNF96_01	RNF96	ZFPHD	1
V$DRI1_01	DRI1	ARID	1
V$LUN1_01	LUN-1	ZFRING	1
V$GMEB2_04	GMEB2	SAND	1
V$REST_Q5	REST	ZFC2H2	1
V$MEF2_03	MEF-2	MADS	1
V$HIF1A_Q5	HIF-1alpha	BHLH	1
V$CIZ_01	CIZ	ZFC2H2	1
V$KAISO_01	Kaiso	ZFC2H2	1
V$CDC5_01	Cdc5	MYB	1
V$BBX_03	Bbx	HMG	2
V$CDX2_01	Cdx-2	HOX	2
V$POU6F1_02	POU6F1	HOX	2
V$HNF4A_Q3	HNF-4A	ZFC4-NR	2
V$CREB1_Q6	CREB1	BZIP	2
V$HSF1_01	HSF1	HSF	2
V$CDX2_Q5_02	CDX-2	HOX	2
V$STAT1_Q6	STAT1	STAT	2
V$RELA_Q6	RelA-p65	REL	2
V$HNF3B_Q6	HNF-3beta	FORKHEAD	2
V$MTF1_Q5	MTF-1	ZFC2H2	3
V$ING4_01	ING4	ZFPHD	3
V$AML3_Q6	AML3	RUNT	3
V$COE1_Q6	COE1	REL	3
V$AIRE_01	AIRE	ZFPHD	3
V$HES1_Q6	HES-1	BHLH	3
V$GFI1_Q6_01	Gfi1	ZFC2H2	3
V$MYB_05	c-Myb	MYB	3
V$CTCF_01	CTCF	ZFC2H2	3
V$ERALPHA_01	ER-alpha	ZFC4-NR	4
V$LRH1_Q5_01	LRH-1	ZFC4-NR	4
V$BLIMP1_Q4	Blimp-1	ZFC2H2	4
V$LEF1_Q5_01	LEF-1	HMG	4
V$SOX10_Q3	Sox10	HMG	4
V$ZNF333_01	ZNF333	ZFC2H2	4
V$TBX5_01	Tbx5	TBX	4
V$MEIS1_01	MEIS1	HOX	4
V$BBX_04	BBX secondary motif	HMG	4
V$POU2F1_Q6	POU2F1	HOX	4
V$FPM315_01	FPM315 (ZNF263)	ZFC2H2	4
V$ZFP105_04	ZFP105 secondary motif	ZFC2H2	4
V$MZF1_Q5	MZF-1	ZFC2H2	4
V$ERALPHA_Q6_01	ER-alpha	ZFC4-NR	5
V$SF1_Q5_01	SF-1	ZFC4-NR	5
V$TATA_01	TATA	TBP	5
V$SP100_04	SP100 secondary motif	SAND	5

Table 4.3 (2/2)

Matrix	Factor	Classification	Number of sites
V$EGR1_Q6	Egr-1	ZFC2H2	5
V$AP1_Q6_02	AP-1	BZIP	5
V$RFX_Q6	RFX	RFX	5
V$XVENT1_01	Xvent-1	HOX	5
V$IPF1_Q5	ipf1	HOX	5
V$SREBP_Q6	SREBP	BHLH	6
V$MYOGENIN_Q6_01	myogenin	BHLH	6
V$FREAC3_01	Freac-3	FORKHEAD	6
V$ZFX_01	Zfx	ZFC2H2	6
V$MAZR_01	MAZR	ZFC2H2	6
V$AHR_Q6	AHR	BHLH	6
V$E2A_Q6_01	E2A	BHLH	6
V$CP2_Q6	CP2	GRAINY	7
V$GATA_Q6	GATA	ZFGATA	7
V$EBOX_Q6_01	Ebox	BHLH	7
V$RFX1_01	RFX1	RFX	8
V$MAF_Q4	MAF	BZIP	8
V$FAC1_01	FAC1	ZFPHD	8
V$ZSCAN4_04	ZSCAN4 secondary motif	ZFC2H2	8
V$DELTAEF1_01	deltaEF1	ZFC2H2	8
V$RBPJK_01	RBP-Jkappa	REL	9
V$HMGIY_Q3	HMGIY	ATHOOK	9
V$ETS_Q6	Ets	ETS	9
V$PBX_Q3	Pbx	HOX	9
V$NF1_Q6	NF-1	SMAD	10
V$TEF1_Q6_04	TEF-1	TEA	11
V$HIC1_08	Hic1	ZFC2H2	11
V$BEN_01	BEN	BHLH	12
V$CREBP1_01	ATF-2	BZIP	13
V$MAZ_Q6_01	MAZ	ZFC2H2	13
V$MYB_Q4	c-Myb	MYB	13
V$TTF1_Q5_01	TTF-1	HOX	14
V$SP1_Q6_01	Sp1	ZFC2H2	14
V$BRCA_01	BRCA1:USF2	ZFRING	15
V$CEBPA_Q6	C/EBPalpha	BZIP	15
V$ZFP161_04	ZF5 secondary motif	ZFC2H2	15
V$CDPCR1_01	CDP CR1	HOX	16
V$CHCH_01	Churchill	CHCH	18
V$MUSCLEINI_B	Muscle initiator	GENINI	18
V$SMAD4_Q6_01	Smad4	SMAD	18
V$RUSH1A_02	RUSH-1alpha	ZFRING	19
V$DR4_Q2	LXR, PXR, CAR, COUP, RAR	ZFC4-NR	23
V$MAFA_Q4	MAFA	BZIP	23
V$P53_Q3	P53	P53	23
V$GLI_Q3	GLI	ZFC2H2	24
V$NF1A_Q6_01	NF-1A	SMAD	24
V$NKX25_Q6	Nkx2.5	HOX	25
V$RREB1_01	RREB-1	ZFC2H2	28
V$GEN_INI_B	GEN_INI	GENINI	30
V$YY1_Q6_03	YY1	ZFC2H2	30
V$CPBP_Q6	CPBP	ZFC2H2	45
V$AP2ALPHA_03	AP-2alphaA	BHSH	52
V$P53_04	p53	P53	53
V$IK_Q5_01	Ikaros	ZFC2H2	61
V$GKLF_Q4	GKLF	ZFC2H2	63

Table 4.3 shows a summary of the identified TFBSs. Similar to the promoters of other genes mentioned above, the only over-represented TFBS in *Foxn1* promoter was BRCA1:USF2. Among the genes analysed in Chapter 3, only *Inhibitor of growth family, member 4* (*Ing4*) had binding sites in *Foxn1* promoter, suggesting that this gene could be involved in regulating *Foxn1* expression. Of particular interest are SMAD4 binding sites identified in this region (Table 4.3). Given the role of SMAD4 as a common mediator-SMAD for both BMP and TGFβ signalling pathways, the above results suggest that these pathways could be involved in regulation of *Foxn1* expression in TECs. As discussed in Chapter 1, both BMP and TGFβ signalling play important roles in thymus development, maintenance, and function and BMP signalling has been suggested to be required for initiation of *Foxn1* expression in TEPCs. Thus, whether TGFβ signalling is also involved in regulation of *Foxn1* expression remains to be determined.

4.7.2 Putative *Foxn1* enhancers

Using the genomic regions assignment criteria mentioned above, four putative enhancer regions, with overlapping H3K4me1 and H3K27ac peaks, were assigned to *Foxn1* (Figure 4.16A). Both the H3K4me1 and H3K27ac tracks showed high signal throughout intron-1 of *Foxn1*, which could result from transcription of the alternate first exon 1b. The genomic coordinates (in mm9/GRCm37) for the predicted enhancer regions assigned to *Foxn1* were as follows:

Region-1: Chr11: 78185221 – 78186298

Region-2: Chr11: 78188296 – 78188610

Region-3: Chr11: 78189231 – 78189789

Region-4: Chr11: 78200392 – 78201601

These regions lie within the 30kb region shown to be able to recapitulate *Foxn1* expression during thymus development (Schlake 2005). A summary of the TFBSs present within these regions is shown in Table 4.4 to Table 4.7. Among the genes analyzed in Chapter 3, *Foxa2*, *Hes1*, *Foxc1*, *E2Fs*, *p53*, and *Ing4* had binding sites in the identified enhancer regions, suggesting their involvement in regulation of *Foxn1*

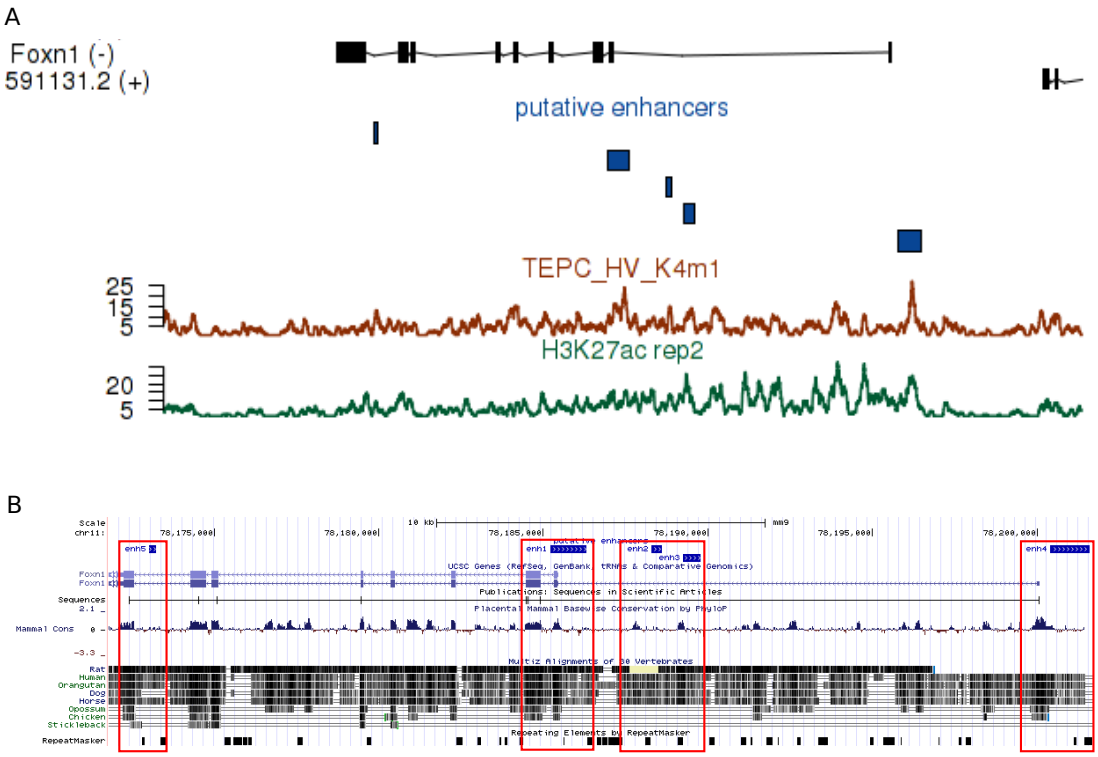


Figure 4.16: Identification of putative *Foxn1* enhancers

(A) Genome browser tracks for H3K4me1 and H3K27ac at *Foxn1* genomic locus. Also shown are identified putative enhancers as determined by overlapping H3K4me1 and H3K27ac peaks. Five putative enhancher regions were identified. (B) UCSC genome browser view of *Foxn1* genomic locus with identified enhancers shown as blue bars (marked by red boxes).

expression. E2Fs have previously been shown to be able regulate the expression of by binding to their consensus sequence in the vicinity of *Foxn1* transcription start site *in vitro* (Garfin et al. 2013). Whether E2Fs and the other genes bind to the identified binding sites in putative enhancer regions remains to be determined. Interestingly, no TFBS was found to be over-represented in any of the predicted enhancer regions. Figure 4.16B shows that the identified putative enhancer regions either overlap or are very close to regions of high conservation, suggesting that these regions could play a role in regulating *Foxn1* transcription. As mentioned previously, FOXN1 appears to regulate the expression of its own gene. However, no FOXN1 binding site was identified in any of the predicted enhancer regions, possibly because of availability of only a poorly characterized binding site matrix for FOXN1.

Besides the above mentioned putative enhancer regions, another region with overlapping H3K4me1 and H3K27ac histone modifications was found between exon-8 and exon-9 of *Foxn1*. This putative enhancer region is, however, not assigned to *Foxn1* using the above mentioned assignment criteria. This is due to the region being closer to the TSS of a neighboring gene. Given that this region is located within *Foxn1* locus, I decided to assign it as the fifth putative enhancer region for *Foxn1*. The genomic coordinates for this region are (in mm9/GRCm37):

Region-5: Chr11: 78173017 – 78173239

The TFBSs identified in this region are listed in Table 4.8. Interestingly, the region between exon-8 and exon-9 of *Foxn1* contained enrichment for H3K4me3, H3K4me1, and H3K27ac in a study that analysed histone modifications in whole adult thymus (Shen et al. 2012). The H3K4me3 data generated here does not show any enrichment in this region, suggesting that H3K4me3 enrichment observed in the above mentioned study results from either the difference between TEPCs and adult TECs or corresponds to another cell-type present in the adult thymus. Region-5 does not contain any over-represented TFBS, nor any binding sites for FOXN1, as determined using Match programme.

Matrix	Factor	Classification	Number of sites
V$COE1_Q6	COE1	REL	2
V$MYOGENIN_Q6_01	myogenin	BHLH	2
V$DMRT4_01	DMRT4	DMRT	1
V$NFY_Q3	NF-Y	HISTONE	1
V$ZFP161_04	ZF5 secondary motif	ZFC2H2	1
V$CRX_Q4_01	CRX	HOX	1
V$RFX1_01	RFX1	RFX	1
V$REST_Q5	REST	ZFC2H2	1
V$ZFX_01	Zfx	ZFC2H2	1
V$XVENT1_01	Xvent-1	HOX	1
V$GMEB2_04	GMEB2	SAND	1
V$HIF1A_Q5	HIF-1alpha	BHLH	1
V$RFX_Q6	RFX	RFX	1
V$LUN1_01	LUN-1	ZFRING	1
V$BLIMP1_Q4	Blimp-1	ZFC2H2	1
V$SRF_Q5_02	SRF	MADS	1
V$HES1_Q6	HES-1	BHLH	1
V$DELTAEF1_01	deltaEF1	ZFC2H2	1
V$CDX2_Q5_02	CDX-2	HOX	1
V$TBX5_01	Tbx5	TBX	1
V$ING4_01	ING4	ZFPHD	1
V$AML3_Q6	AML3	RUNT	1

Table 4.4: Summary of transcription factor binding sites identified in *Foxn1* enhancer-1
Foxn1 enhancer-1 was analysed for presence of transcription factor binding sites using Match algorithm and TRANSFAC professional database. This table lists the matrices, the associated transcription factors, and number of sites identified for each matrix.

Matrix	Factor	Classificaiton	Number of sites
V$GEN_INI_B	GEN_INI	GENINI	6
V$AP2ALPHA_03	AP-2alphaA	BHSH	4
V$IK_Q5_01	Ikaros	ZFC2H2	3
V$P53_04	p53	P53	2
V$LEF1_Q5_01	LEF-1	HMG	2
V$DR4_Q2	LXR, PXR, CAR, COUP, RAR	ZFC4-NR	2
V$RELA_Q6	RelA-p65	REL	2
V$MYOGENIN_Q6_01	myogenin	BHLH	2
V$CEBPA_Q6	C/EBPalpha	BZIP	2
V$YY1_Q6_03	YY1	ZFC2H2	2
V$BRCA_01	BRCA1:USF2	ZFRING	2
V$E2A_Q6_01	E2A	BHLH	2
V$RFX1_01	RFX1	RFX	2
V$GLI_Q3	GLI	ZFC2H2	1
V$ERALPHA_Q6_01	ER-alpha	ZFC4-NR	1
V$HES1_Q6	HES-1	BHLH	1
V$RREB1_01	RREB-1	ZFC2H2	1
V$IPF1_Q5	ipf1	HOX	1
V$SF1_Q5_01	SF-1	ZFC4-NR	1
V$CHCH_01	Churchill	CHCH	1
V$ETS_Q6	Ets	ETS	1
V$CREB1_Q6	CREB1	BZIP	1
V$PBX_Q3	Pbx	HOX	1
V$LRH1_Q5_01	LRH-1	ZFC4-NR	1
V$CP2_Q6	CP2	GRAINY	1
V$SMAD4_Q6_01	Smad4	SMAD	1
V$GATA_Q6	GATA	ZFGATA	1
V$TBX5_01	Tbx5	TBX	1
V$ZFP105_04	ZFP105 secondary motif	ZFC2H2	1
V$MEIS1_01	MEIS1	HOX	1
V$BEN_01	BEN	BHLH	1
V$AP1_Q6_02	AP-1	BZIP	1
V$NF1_Q6	NF-1	SMAD	1
V$GKLF_Q4	GKLF	ZFC2H2	1
V$P53_Q3	P53	P53	1
V$SRY_Q6	SRY	HMG	1
V$TATA_01	TATA	TBP	1
V$POU2F1_Q6	POU2F1	HOX	1
V$NF1A_Q6_01	NF-1A	SMAD	1
V$FAC1_01	FAC1	ZFPHD	1
V$CDPCR1_01	CDP CR1	HOX	1
V$MYB_Q4	c-Myb	MYB	1
V$HNF3B_Q6	HNF-3beta	FORKHEAD	1

Table 4.5: Summary of transcription factor binding sites identified in *Foxn1* enhancer-2
Foxn1 enhancer-2 was analysed for presence of transcription factor binding sites using Match algorithm and TRANSFAC professional database. This table lists the matrices, the associated transcription factors, and number of sites identified for each matrix.

Matrix	Factor	Classificaiton	Number of sites
V$XVENT1_01	Xvent-1	HOX	1
V$BRCA_01	BRCA1:USF2	ZFRING	1
V$GCM2_01	GCMb	GCM	1
V$DRI1_01	DRI1	ARID	1
V$FPM315_01	FPM315 (ZNF263)	ZFC2H2	1
V$SRY_Q6	SRY	HMG	1
V$SREBP_Q6	SREBP	BHLH	1
V$BBX_04	BBX secondary motif	HMG	1
V$EKLF_Q5_01	EKLF	ZFC2H2	1
V$RBPJK_01	RBP-Jkappa	REL	1
V$P53_Q3	P53	P53	1
V$MZF1_Q5	MZF-1	ZFC2H2	1
V$GFI1_Q6_01	Gfi1	ZFC2H2	1
V$SOX10_Q3	Sox10	HMG	1

Table 4.6: Summary of transcription factor binding sites identified in *Foxn1* enhancer-3
Foxn1 enhancer-3 was analysed for presence of transcription factor binding sites using Match algorithm and TRANSFAC professional database. This table lists the matrices, the associated transcription factors, and number of sites identified for each matrix.

Matrix	Factor	Classification	Number of sites
V$RFX1_01	RFX1	RFX	2
V$BRCA_01	BRCA1:USF2	ZFRING	2
V$FREAC3_01	Freac-3	FORKHEAD	2
V$NF1_Q6	NF-1	SMAD	2
V$BBX_03	Bbx	HMG	2
V$EGR1_Q6	Egr-1	ZFC2H2	1
V$HMGIY_Q3	HMGIY	ATHOOK	1
V$TATA_01	TATA	TBP	1
V$PIT1_Q6_01	Pit-1	HOX	1
V$LUN1_01	LUN-1	ZFRING	1
V$REST_01	REST	ZFC2H2	1
V$LRH1_Q5_01	LRH-1	ZFC4-NR	1
V$SRY_Q6	SRY	HMG	1
V$MAFA_Q4	MAFA	BZIP	1
V$RORALPHA_Q4	RORalpha	ZFC4-NR	1
V$SOX10_Q3	Sox10	HMG	1
V$MTF1_Q5	MTF-1	ZFC2H2	1
V$ZFP105_04	ZFP105 secondary motif	ZFC2H2	1
V$CDC5_01	Cdc5	MYB	1
V$ZNF333_01	ZNF333	ZFC2H2	1
V$CHCH_01	Churchill	CHCH	1
V$SF1_Q5_01	SF-1	ZFC4-NR	1
V$NFY_Q3	NF-Y	HISTONE	1
V$COE1_Q6	COE1	REL	1
V$SOX2_Q3_01	Sox2	HMG	1
V$CDX2_01	Cdx-2	HOX	1
V$STAT1_Q6	STAT1	STAT	1
V$BBX_04	BBX secondary motif	HMG	1
V$MZF1_Q5	MZF-1	ZFC2H2	1
V$HIC1_08	Hic1	ZFC2H2	1
V$ISL1_Q3	islet1	HOX	1
V$ING4_01	ING4	ZFPHD	1
V$DELTAEF1_01	deltaEF1	ZFC2H2	1
V$POU2F1_Q6	POU2F1	HOX	1

Table 4.7: Summary of transcription factor binding sites identified in *Foxn1* enhancer-4
Foxn1 enhancer-4 was analysed for presence of transcription factor binding sites using Match algorithm and TRANSFAC professional database. This table lists the matrices, the associated transcription factors, and number of sites identified for each matrix.

Matrix	Factor	Classification	Number of sites
V$GKLF_Q4	GKLF	ZFC2H2	6
V$IK_Q5_01	Ikaros	ZFC2H2	5
V$MAZ_Q6_01	MAZ	ZFC2H2	3
V$TTF1_Q5_01	TTF-1	HOX	3
V$SP100_04	SP100 secondary motif	SAND	3
V$CEBPA_Q6	C/EBPalpha	BZIP	2
V$HES1_Q6	HES-1	BHLH	2
V$RBPJK_01	RBP-Jkappa	REL	2
V$GEN_INI_B	GEN_INI	GENINI	2
V$SP1_Q6_01	Sp1	ZFC2H2	2
V$XVENT1_01	Xvent-1	HOX	2
V$CPBP_Q6	CPBP	ZFC2H2	2
V$NF1A_Q6_01	NF-1A	SMAD	1
V$E2F_Q6_01	E2F	E2F	1
V$ZFP161_04	ZF5 secondary motif	ZFC2H2	1
V$ETS_Q6	Ets	ETS	1
V$CHCH_01	Churchill	CHCH	1
V$RFX1_01	RFX1	RFX	1
V$POU6F1_02	POU6F1	HOX	1
V$NKX25_Q6	Nkx2.5	HOX	1
V$ERALPHA_01	ER-alpha	ZFC4-NR	1
V$CP2_Q6	CP2	GRAINY	1
V$SMAD4_Q6_01	Smad4	SMAD	1
V$EGR1_Q6	Egr-1	ZFC2H2	1
V$LEF1_Q5_01	LEF-1	HMG	1
V$INSM1_01	INSM1	CHCH	1
V$RREB1_01	RREB-1	ZFC2H2	1
V$GLI_Q3	GLI	ZFC2H2	1
V$SRY_Q6	SRY	HMG	1
V$SREBP_Q6	SREBP	BHLH	1
V$MAF_Q4	MAF	BZIP	1
V$CDPCR1_01	CDP CR1	HOX	1
V$EBOX_Q6_01	Ebox	BHLH	1
V$FPM315_01	FPM315 (ZNF263)	ZFC2H2	1
V$AP1_Q6_02	AP-1	BZIP	1
V$HNF3B_Q6	HNF-3beta	FORKHEAD	1
V$MEIS1_01	MEIS1	HOX	1
V$ERALPHA_Q6_01	ER-alpha	ZFC4-NR	1

Table 4.8: Summary of transcription factor binding sites identified in *Foxn1* enhancer-5
Foxn1 enhancer-5 was analysed for presence of transcription factor binding sites using Match algorithm and TRANSFAC professional database. This table lists the matrices, the associated transcription factors, and number of sites identified for each matrix.

4.8 Discussion

The aim of the work described in this chapter was to identify putative transcriptional regulatory regions in TEPCs. To this end, ChIP-seq data was generated for H3K4me3, H3K4me1, and H3K27ac from isolated TEPCs, permitting identification of active promoters and putative active enhancers. This chapter also discusses the difficulties and limitations associated with this approach. One of the key limitations was the availability of material for ChIP. Isolated TEPCs were used for ChIP in order to ensure that the results closely match the actual biology of the developing thymus. An alternate source of material for ChIP would be thymic epithelial cell lines, however TEC are known to lose key features of their *in-vivo* phenotype extremely quickly upon establishment of primary cultures, indicating that analysis of TEC lines is of limited use. Furthermore, several protocols have been published on differentiation of pluripotent stem cells into thymic epithelial-like cells but so far bona fide TEPC have not been generated (Bredenkamp, Ulyanchenko, et al. 2014). Thus, the use of isolated TEPCs was the best approach for the questions I set out to address.

The advancements in ChIP-seq protocols now enable this technique to be performed on much smaller number of cells, as described here. This has led to this technique being applied to various tissues and organs in the last few years (Adli et al. 2010; Adli & Bernstein 2011). Here, I demonstrate, for the first time, the application of this technique for the identification of genome wide transcriptional regulatory regions in TEPCs. Of note is that ChIP-seq for histone modifications has previously been applied to whole adult thymus (Shen et al. 2012) and to mature mTECs (Sansom et al. 2014). However, the presence of other cell types in the adult thymus, which constitutes the majority of adult thymus cellularity, confounds the results presented in the former study, preventing identification of histone modifications relevant to thymic epithelial cells. This is evident from an absence of H3K4me3 enrichment at *Foxn1* TSS in this dataset. However, as mentioned previously, the enrichment for H3K4me1 and H3K27ac observed towards the 3` end of *Foxn1* might represent a TEC specific enrichment. On the other hand, the later study by Sansom et al. only

analysed H3K4me3 and H3K27me3 histone modifications in mTECs and thus does not enable identification of active enhancers. Furthermore, mTECs express a large number of tissue-restricted antigens (TRAs), that are important for negative selection of developing thymocytes, and are thus expected to present histone modifications at promoters and enhancers for these genes. This further complicates identification of histone modifications relevant to the aim of this Chapter. The results described in this Chapter overcome both these issues and thus represent a significant advancement to our understanding of TEPC epigenetics.

4.8.1 The effect of NaBu on ChIP

NaBu was added to all rep-2 samples used for ChIP. Whether the above-mentioned differences observed between rep-1 and rep-2 are a result of NaBu or of technical variation cannot be confidently determined without the use of more biological replicates. Furthermore, whether NaBu is required during the entire isolation, sorting, and subsequent ChIP protocol, as done here, is not clear. Comparison of samples treated at different stages (isolation, sorting, or ChIP) with NaBu could help resolve this issue. However, this requires the use of more animals for optimization and was thus avoided in the present study.

4.8.2 Identified putative transcriptional regulatory regions are important for genes involved in thymus development

The ChIP-seq approach employed here allowed for the identification of genome-wide promoters and enhancer elements. This enables testing of various genetic interactions prevalent within a particular cell type. Determining the exact function of identified genomic regions, however, requires genetic manipulations. Here, I carried out preliminary analysis of the identified putative enhancer regions by determining ontology enrichment for genes in the vicinity of these regions, as per the default parameter in GREAT bioinformatics package. This analysis showed that many of the identified putative enhancer regions were in the vicinity of genes involved in thymus development. Changing the genomic region assignment parameters such that each

region was only assigned to its nearest gene changed the list of enriched terms. The "thymic hypoplasia" term was replaced by "abnormal thymus morphology" suggesting that the observed enrichment for genes important in thymus was not an artefact of assigning regions to genes. However, these two phenotype terms also include genes involved in T-cell development defects, which result in abnormal thymus development. Thus, the role of these genes in TECs remains to be determined experimentally.

Of particular interest was the enrichment for TGFβ signalling pathway. Analysing the genes leading to this enrichment revealed genes specific to TGFβ pathway and not the BMP, Nodal, or Activin pathway, which are also members of transforming growth factor superfamily. This suggests an important role for TGFβ signalling pathway in TEPCs. Several studies have demonstrated the importance of TGFβ in postnatal thymus maintenance and function (Sempowski et al. 2000; Kumar et al. 2006; Hauri-Hohl et al. 2008; Odaka et al. 2013; Hauri-Hohl et al. 2014). Thus, its role, if any, during thymus development warrants further study. TGFβ signals through various downstream effectors of which SMAD- and p38-dependant pathways were highly enriched. SMAD4 binding sites were found to be present within the promoter regions of various genes important for thymus development, suggesting that SMAD4 is important for this process. Some of the other pathways enriched in the present analysis included the BMP, Notch, and Wnt signalling pathways. It is well established that BMP signalling is important for thymus development and normal *Foxn1* expression (Bleul & Boehm 2005; Patel et al. 2006; Gordon et al. 2010; Neves et al. 2012). Furthermore, the Wnt signalling pathway has been suggested to be important for TEPCs/TECs (Huh & Ornitz 2010; Bredenkamp, Nowell, et al. 2014), and the Notch signalling is currently being investigated in our lab for its role in thymus development and maintenance.

4.8.3 Additional techniques for interogating chromatin accessibility and enhancers

While the histone modification ChIP-seq technique used here allowed identification of putative global enhancers, identification of p300 binding sites and/or open chromatin structures in TEPCs can further complement the data presented in this Chapter. The transcriptional coactivator and acetyltransferase p300, which binds to enhancers, has been used extensively for identification of genome-wide enhancers (Rada-Iglesias et al. 2011). However, given its transient interaction with DNA, identification of p300 binding sites through ChIP-seq on small number of cells, such as those used in this Chapter, presents additional challenges. On the other hand, the development of ATAC-seq protocol now allows studying chromatin accessibility without the need of millions of cells (Buenrostro et al. 2013). Combining chromatin accessibility data together with the histone modification data presented in this Chapter could further filter for regions that are most likely to act as active enhancers. Thus, ATAC-seq and p300 ChIP-seq could be performed on TEPCs to further support the identification of enhancers.

4.8.4 Concluding remarks

The importance of *Foxn1* in thymus has been realised since its discovery as the *nude* gene. Understanding its regulation and function in TEPCs and TECs has become even more important in recent times since discoveries of its involvement in regenerating the thymus *in-vivo* and to reprogramme MEFs to iTECs (Bredenkamp, Nowell, et al. 2014; Bredenkamp, Ulyanchenko, et al. 2014). Here, I have identified genomic regions in the vicinity of *Foxn1* locus marked by H3K4me3, H3K4me1, and H3K27ac histone modifications. The identified promoter and putative enhancers, except for Region-5, lie within the 30kb region shown to be able to recapitulate *Foxn1* expression during thymus development. Whether the regions identified as putative enhancers are able to recapitulate this expression pattern remains to be determined. Interestingly, Region-5 was also detected by Shen et al. in adult thymus and could thus represent an enhancer important for the maintenance of *Foxn1* expression. Similarly, whether the region identified as the promoter is sufficient for

basal transcription remains to be determined. The luciferase assay can be employed to test the promoter and enhancer functions of these regions. However, this assay only addresses the function of these regions *in-vitro*, which could be different from their activity *in-vivo*. Further validation for enrichment of H3K4me1 and H3K27ac should be carried out using ChIP-qPCR. Comparison of enrichment of these histone modifications at putative enhancer regions to that in TEPCs isolated from E10.5-E11.0, i.e. before the initiation of *Foxn1* transcription, should be carried out to test whether these regions behave as true enhancers.

The advance in genome engineering resulting from the discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 as genome-editing tool provides a unique opportunity to test the *in-vivo* functionalities of regions identified through such ChIP-seq approaches. Thus, deletion of the identified putative enhancers, singly and in pairs, using CRISPR-Cas9 system should be performed to determine their roles in regulating *Foxn1* expression. As shown in the tables above, the identified promoter and enhancer regions each contain binding sites for many transcription factors. Whether all of these transcription factors are expressed in TEPCs is not yet known. The mentioned CRISPR-Cas9 system can also be used to either delete or mutate the binding sites for these transcription factors to study their role in regulating *Foxn1* transcription.

5. Transcriptional regulation of TEPC development and differentiation

5.1 Introduction

In Chapter 3, I described identification of a potential transcription factor network governing TEPC differentiation, while in Chapter 4, I identified, on a global basis, putative transcriptional regulatory regions active in TEPC. From this information, I predicted signalling pathways and transcription factors with potential roles in TEPC regulation. However, while Chapter 4 identified transcription factors with physical binding sites within transcriptional regulatory regions of *Foxn1*, with the exception of the factors analysed in Chapter 3, very little is known about their expression during thymus development. The same is true for the genes involved in the signalling pathways predicted to be important for TEPC in Chapter 4. This Chapter therefore describes analysis of the transcriptome of developing 3PP cells and TEPCs using RNA-seq, which was performed in order to resolve these issues. The 3PP undergoes several morphological changes during development into the thymic rudiment, which are likely to be accompanied by a change in transcriptional programme from one that is important for endoderm development and/or tissue/organ development to one that is required for TEPC specification and differentiation. To test this hypothesis and to determine the genes involved in these processes, I performed enrichment analysis on the RNA-seq data using bioinformatics approaches. The outcomes of these analyses are reported below.

5.2 Results

5.2.1 Sample preparation

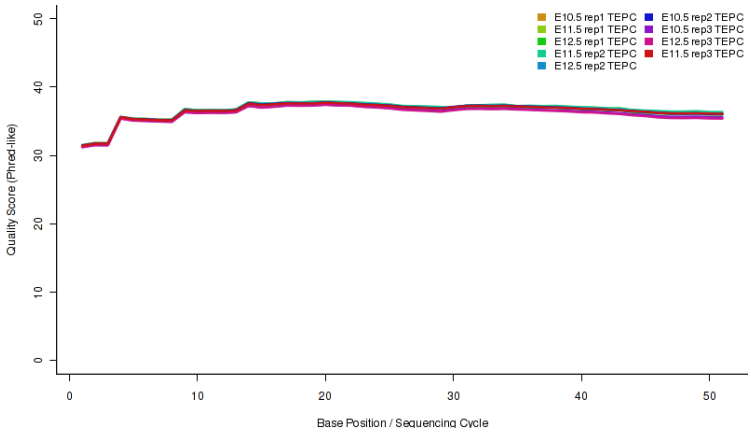
Advances in the field of transcriptomics, particularly RNA-seq, have now made it possible to characterize transcriptomes of even difficult to obtain cell types, such as adult stem and progenitor cells. The ability to obtain global transcriptome data from very small number of cells (even single cells) using advanced chemistries such as

SMART-seq2 (Picelli et al. 2014) has opened up the possibility of characterising the transcriptome of defined subsets of thymic epithelial cells (TEC). Thus, I isolated 100 TEPCs from E10.5, E11.5, and E12.5 thymic primordial, using flow cytometry as described in Chapter 3, and processed the cells for RNA-seq. Three independent biological replicates were collected for each sample, in order to allow statistical interpretation of the resulting data. The selected developmental time points cover both *Foxn1*-independent development and *Foxn1*-dependent differentiation of TEPCs. This work was carried out in collaboration with Prof. Sten Eirik Jacobsen's laboratory at the Karolinska Institute and the University of Oxford. Cell sorting was performed at the University of Edinburgh (by myself). Preparation of these sorted samples for RNA-seq and subsequent sample preparation and sequencing were carried out at University of Oxford using Smart-seq2 protocol (Picelli et al. 2014). The Smart-seq2 protocol is optimized for use with very small cell numbers, even single cells and demonstrates better sensitivity, accuracy, and full-length transcript coverage than other protocols (Picelli et al. 2014). The current limitations of this protocol are that it does not retain strand specificity of RNA samples and is unable to detect nonpolyadenylated RNA (Picelli et al. 2014).

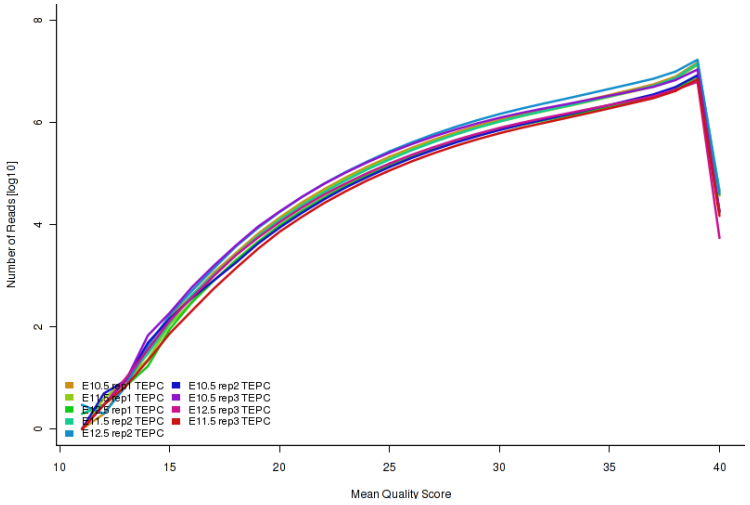
5.2.2 Comparison of biological replicates

The sequenced reads generated from each sample were obtained from Oxford in fastq format. These files were uploaded to GeneProf (Halbritter et al. 2011) for quality control assessment of the sequenced reads and their alignment to transcripts using the TopHat 1.2.0 algorithm (Trapnell et al. 2009). The raw read counts (RC) and read counts normalized for library sizes and gene lengths (RPKM; reads per kilo-base of exon model (entire annotated transcript) per million mapped reads) were generated, subsequent to reads alignment, using the "Quantitate Gene Expression" module of GeneProf (Halbritter et al. 2011). The global RC and RPKM values were downloaded and used for further analysis. The sequenced reads were of high quality (Figure 5.1A-B) and between 55 and 65% of all aligned reads for each sample were aligned uniquely in the genome (Figure 5.1C), suggesting that the samples used for generating sequencing libraries were of good quality.

A: Quality score per base position



B: Distribution of mean read quality scores



C: Distribution of number of matches in genome for sequenced reads

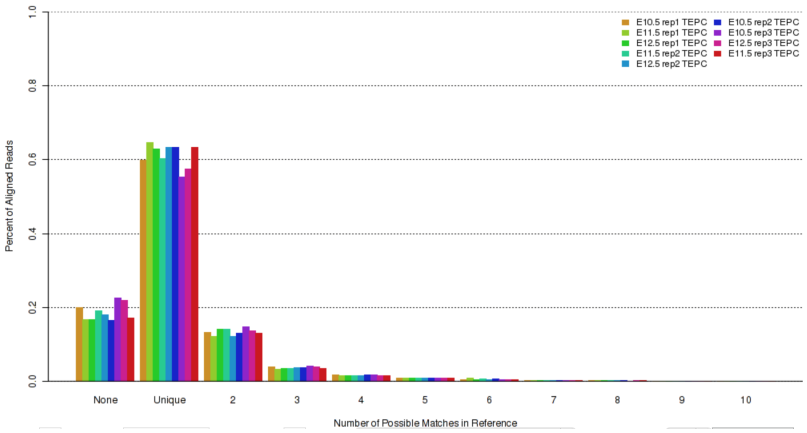


Figure 5.1: Quality control of RNA-seq data

(A) A plot of quality score per sequencing cycle averaged for all the sequenced reads for each RNA-seq sample. (B) A plot showing the distribution of average quality scores for sequenced reads for each RNA-seq sample.The quality scores shown represent Phred-like scores which measures the probability that a base is called incorrectly. Higher scores represent a smaller probability of error. (C) Distribution of number of matched in genome for sequenced reads for each RNA-seq sample. Note that most of the sequenced reads aligned to a single location in the genome.

I first tested the reproducibility of the RNA-seq data generated from each replicate by calculating the correlation coefficient for RPKM counts between biological replicates (Figure 5.2A-C). This ranged from 0.88 to 0.98, indicating a very high degree of similarity between normalised read counts among biological replicates. The most commonly used diagnostic plots for RNA-seq data include principal component analysis (PCA) and sample clustering. Therefore, the matrix of raw read count obtained from GeneProf was converted into a DESeqDataSet object using the DESeq2 package (Anders & Huber 2010; Love et al. 2014) in the statistical computing language R (Core Team 2015). The raw read counts were subsequently normalized using the regularized logarithm transformation (rlog) approach of DESeq2, which prevents the results being dominated by genes on either end of the expression range and has been shown to perform equally well or better than other normalisation approaches (Love et al. 2014). Analysis of variance in rlog-transformed data through PCA showed that the majority (75%) of variance could be accounted for by two principal components (Figure 5.3A). Furthermore, the PCA analysis revealed that based on these two principal components, the biological replicates for the E12.5 and E11.5 samples were highly similar. On the other hand, E10.5 rep2 appeared more similar to the E11.5 samples than to E10.5 rep1 and rep3 (Figure 5.3A). Similarly, a heatmap of rlog-transformed samples, generated using the pheatmap function (Kolde 2015), clustered E10.5 rep2 closer to the E11.5 samples than to the other two E10.5 replicates (Figure 5.3B). Thus, the E10.5 rep2 sample appeared to be of a developmental time point between E10.5 and E11.5 and was therefore not included in subsequent analyses. The heatmap also showed a high degree of similarity between the biological replicates of both the E12.5 and E11.5 samples, consistent with the PCA results. Thus, all the biological replicates for both E12.5 and E11.5 were suitable for downstream analysis. Finally, consistent with previous suggestions (Nowell et al. 2011), the E11.5 samples clustered closer with the E10.5 samples than with the E12.5 samples, suggesting that the global transcriptome profile of E11.5 3PP cells is more similar to E10.5 3PP cells than to E12.5 TEPCs.

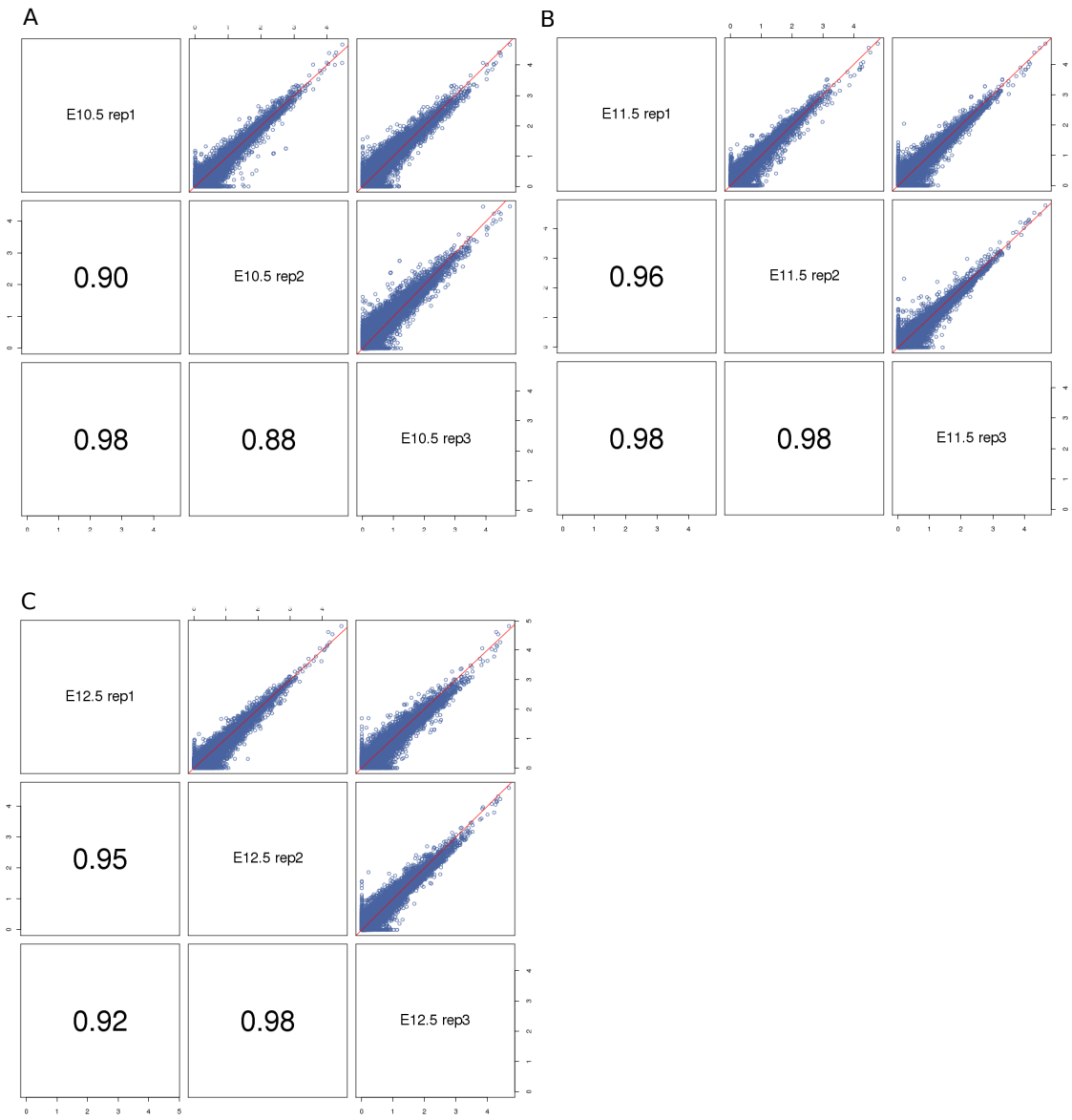


Figure 5.2: Correlation analysis between biological replicates for RNA-seq samples
Correlation coefficients between biological replicates for RNA-seq samples were calculated as one measure of reproducibility between replicates. Shown are scatter plot demonstrating correlation between RPKM values and respective correlation coefficients between each pair of biological replicates for (A) E10.5 and (B) E11.5 third phayrngeal pouch endoderm cells and (C) E12.5 thymic epithelial progenitor cells.

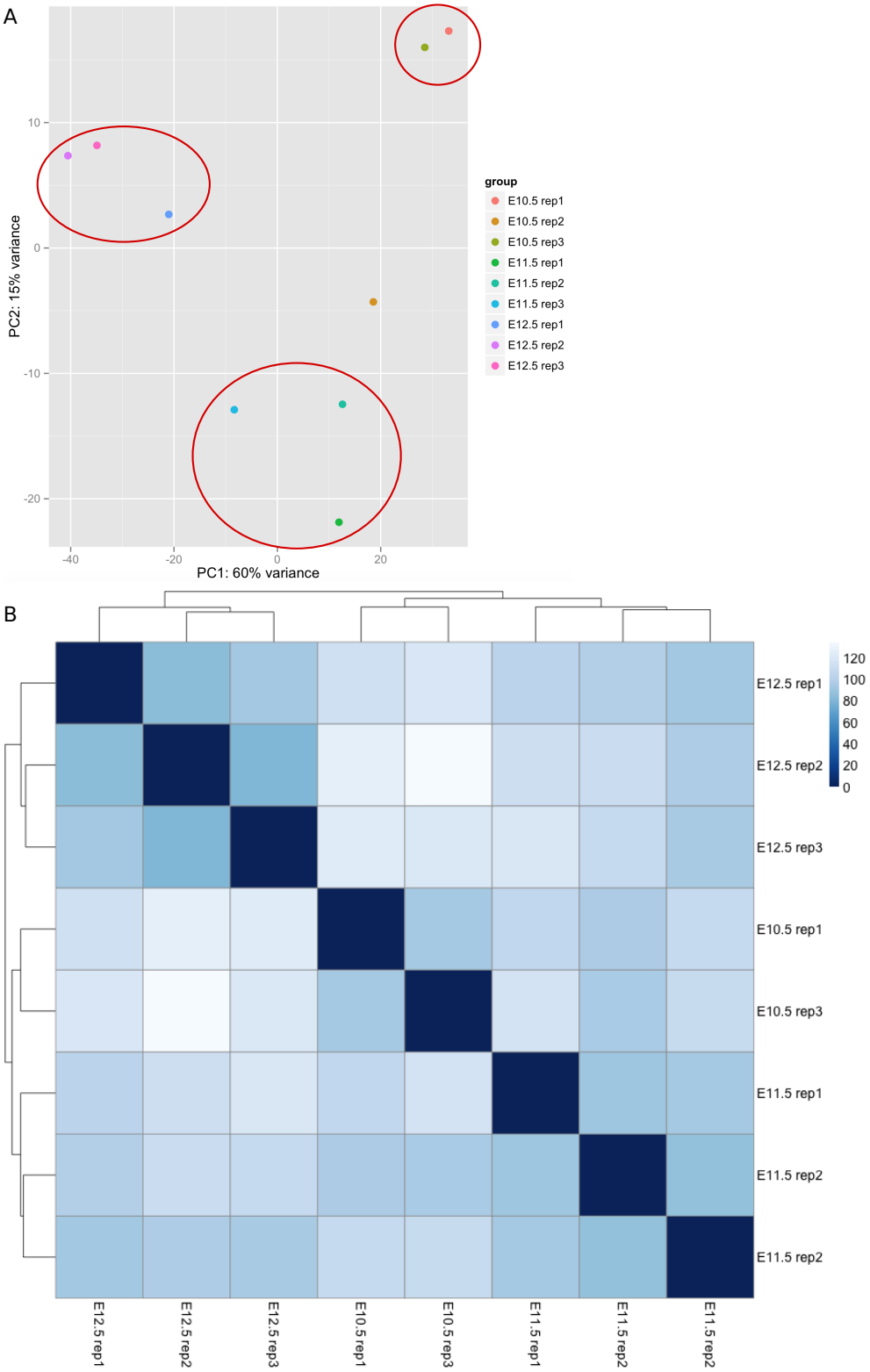


Figure 5.3: Principal component and clustering analysis of RNA-seq biological replicates
(A) Principal component analysis based on rlog-transformed values for all RNA-seq samples. Note the clustering of samples based on developmental time point, except E10.5 rep-2. (B) Heatmap demonstrating the similarities between biological replicates based on clustering of samples. Clusters were genereated using rlog-transformed values. Note that E10.5 rep-2 does not cluster with other E10.5 samples.

5.2.3 The top 500 most variable genes identify transcriptional changes associated with TEPC generation and differentiation

Patterns of changes in gene expression can provide useful insight into changes associated with experimental variables - in this case, development from E10.5 3PP cells to E12.5 TEPCs. I therefore decided to analyse patterns of gene expression in the transcriptome data generated here. A closer inspection of the PCA plot shown in Figure 5.3 suggested that the top 500 most variable genes contribute to most of the variance in the data. Thus, I identified the top 500 genes with the highest variability between the samples based on rlog-transformed data, and clustered these genes on a heatmap using the pheatmap function (Figure 5.4). The heatmap shows the deviation of genes in each sample from the average expression across all samples. Genes were divided into three clusters depending on their expression levels: **1**) genes with transient increase in expression at E1.5; **2**) genes with decreasing expression from E10.5 to E12.5; and **3**) genes with increasing expression from E10.5 to E12.5. The biological processes GO terms enriched for each cluster, at the Benjamini value cut-off (FDR) of 0.1, as determined by the DAVID tool (Huang et al. 2008; Huang et al. 2009) are shown in Figure 5.4.

Cluster 1 showed enrichment for genes involved in the GO term "positive regulation of biosynthetic process" (p-value = 0.001). Cluster 2 is enriched for genes involved in biological processes such as "cell fate commitment", "embryonic morphogenesis", "sensory organ development", "cell migration", "embryonic organ development" and other development related terms. Finally, Cluster 3 is enriched for genes involved in "immune response" and "antigen processing and presentation of exogenous peptide antigens", functions performed by differentiated, functional TECs. Thus, this cluster identifies genes that are important for differentiation of TEPCs and/or the normal function of the differentiating TECs. These results suggests that there indeed is a decrease in the expression of genes associated with 3PP development and a subsequent increase in the expression of genes important for thymus development and/or function.

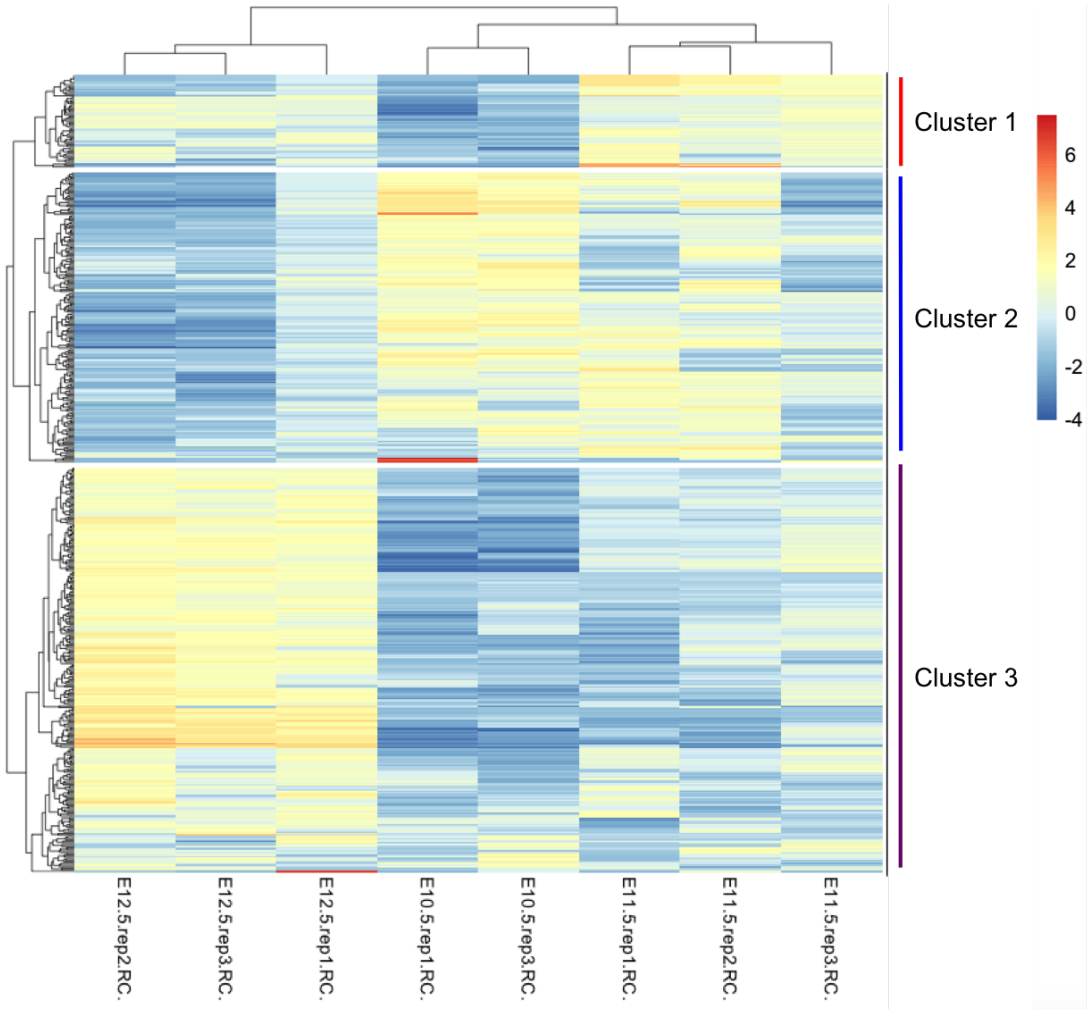


Figure 5.4: Heatmap of top 500 most variable genes.
The top 500 genes with most variablibility in expression between samples, and contributing to most of the variance observed in PCA plot shown previously, were identified using DESeq2. Shown here is a heatmap clustering these genes into three main clusters based on thier expression profiles. The biological processes terms enriched for each cluster are as follows - Cluster-1: nill ; Cluster-2: cell fate commitment, embryonic morphogenesis, sensory organ development, cell migration, embryonic organ development, etc.; Cluster-3: immune response, antigen processing and presentation of exogenous peptide antigens

To ensure that increased expression of genes in Cluster 3 at E12.5 is not a result of presence of contaminating thymocytes in the samples, I analysed the expression of *CD45*, which is expressed by the developing thymocytes but not TECs. The RPKM values for *CD45* was zero in all samples, indicating an absence of expression of this gene and suggesting that samples did not contain any thymocytes.

Nkx2.6 has shown to be expressed widely across the 3PP at E9.0 and E10.0 (Wei & Condie 2011). At E10.0, *Nkx2.6* is expressed throughout the ventral part of 3PP, overlapping the expression of *Isl1*, another gene important for thymus development. My analysis showed that the expression of *Nkx2.6* was decreased from E10.5 to E11.5 (although this was not statistically significant), while its expression was absent at E12.5. Contrary to this, a previous report suggested the absence of *Nkx2.6* expression in E11.5 3PP, based on *in-situ* hybridization (Wei & Condie 2011). This difference might be due to the higher sensitivity of RNA-seq technique employed here compared to *in-situ* hybridization. Whether the decrease in expression of *Nkx2.6* from E10.5 to E11.5 and its subsequent absence at E12.5 is necessary for normal differentiation of TEPCs remains to be determined. No binding sites were found for *Nkx2.6* in the *Foxn1* promoter and enhancers identified in Chapter 4. Finally, the role of other transcription factors in Cluster 4 in 3PP formation and/or thymus development is not yet known.

Cluster 3 represents genes whose expression increases from E10.5 to E12.5. Thus, these genes could represent transcriptional regulators or targets of *Foxn1*. Consistent with this hypothesis, this cluster contains genes such as Dll4, *Ccl25*, *Kitl* and *H2-Ab1*, whose expression in TEPC is *Foxn1*-dependent. *Bhlhe40*, a gene analysed in Chapter 3 as candidate transcriptional regulator of *Foxn1* is also present in this cluster. Finally, *Il7* a gene whose expression is independent of *Foxn1* but crucial for thymocyte development is also present in this cluster. Together, this suggest that Cluster 3 contains genes which are likely to be important for thymus development and may be involved in the same genetic network as *Foxn1*.

Thus, the transcriptome data generated using RNA-seq revealed discrete gene expression patterns that distinguished E10.5 and E11.5 3PP cells from TEPCs. Furthermore, the clusters in the heatmap identify genes that could have important roles in either development/fate commitment of 3PP cells or the differentiation and/or function of TECs.

5.2.4 Analysis of differential gene expression in 3PP cells and TEPC from E10.5 to E12.5

To gain further information regarding genes that were differentially expressed between the E10.5 3PP, E11.5 3PP and E12.5 TEPC samples, I used DESeq2 (Love et al. 2014), an improved version of the DESeq algorithm. DESeq2 identifies differentially expressed genes and provides information on fold change and its statistical significance (Anders & Huber 2010). The DESeqDataSet object generated above was used for these analyses. Pairwise differential gene expression analysis was carried out for all pairs of the three different samples used for RNA-seq. A summary of the results from the DESeq2 analysis is shown in Figure 5.5A.

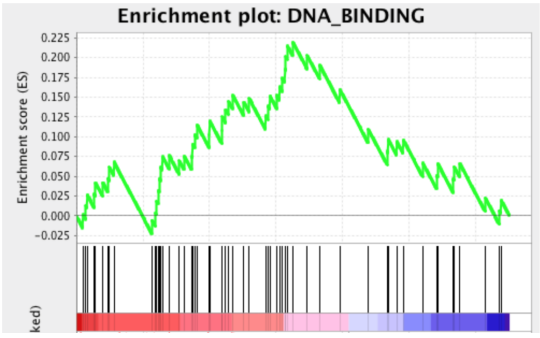
Six hundred and eighty five, nine hundred, and one thousand six hundred and thirty genes were differentially expressed in the E11.5 vs E10.5, E12.5 vs E11.5, and E12.5 vs E10.5 comparisons respectively, using fold change ≥ 2 and p-value ≤ 0.01 as the cut-off at a 10% FDR (false discovery rate). For the E11.5 vs E10.5 comparison, 488 genes were up-regulated at E11.5, while 197 genes were down-regulated. In the E12.5 vs E11.5 comparison, 480 and 421 genes were up- and down-regulated respectively at E12.5. Finally, for the comparison between E12.5 and E10.5, the two

A: Number of differentially expressed genes in each comparison

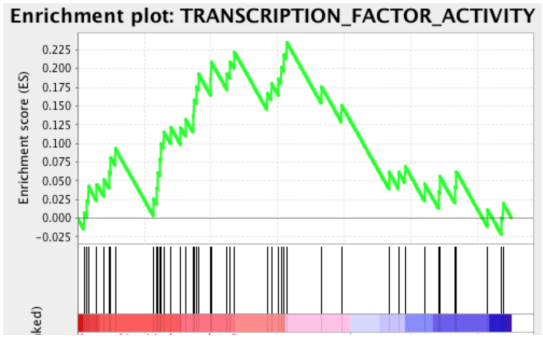
Comparison	# differentially expressed genes	# up-regulated	# down-regulated
E11.5 vs E10.5	685	488	197
E12.5 vs E11.5	900	480	420
E12.5 vs E10.5	1630	1026	604

B: Molecular functional terms enrichment at E12.5 (vs E10.5)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val
1	DNA_BINDING	Details ...	59	0.22	1.99	0.004	0.136
2	TRANSCRIPTION_FACTOR_ACTIVITY	Details ...	45	0.24	1.86	0.006	0.146
3	TRANSCRIPTION_ACTIVATOR_ACTIVITY	Details ...	16	0.30	1.44	0.082	0.826
4	SUBSTRATE_SPECIFIC_TRANSPORTER_ACTIVITY	Details ...	37	0.20	1.41	0.098	0.709
5	SUBSTRATE_SPECIFIC_CHANNEL_ACTIVITY	Details ...	16	0.29	1.36	0.138	0.712



Genes contributing to enrichment of DNA Binding:
Egr2, Nfatc2, Hoxa7, Kcnip3, Pou6f1, Egr1, Fosl1, Stat4, Zbtb38, Nfkb2, Foxo3, Atf3, Stat1, Nfatc1, Sphk1, Tcf19, Trim29, Nfkb1, Mnt, Relb, Jun, Dmrt2, Hic1, Hey1, Fosl2, Bcl6, Foxo1, Irf9, Stat3, Zhx2, Ppargc1a, Arid5b, Nfib, Rad50, Myc, Rscan1, Hes6, Foxg1, Srebf1, Xrcc5, Hes1, Cebpg



Genes contributing to enrichment of Transcription Factor Activity:
Egr2, Nfatc2, Hoxa7, Pou6f1, Egr1, Fosl1, Stat4, Zbtb38, Nfkb2, Foxo3, Atf3, Stat1, Nfatc1, Tcf19, Trim29, Nfkb1, Mnt, Relb, Jun, Dmrt2, Hic1, Hey1, Fosl2, Irf9, Stat3, Zhx2, Nfib, Myc, Rscan1, Hes6, Foxg1, Srebf1

C: Molecular functional terms enrichment at E10.5 (vs E12.5)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val
1	RECEPTOR_ACTIVITY	Details ...	44	-0.30	-2.32	0.004	0.009
2	TRANSMEMBRANE_RECEPTOR_ACTIVITY	Details ...	32	-0.29	-1.92	0.008	0.059
3	PROTEIN_KINASE_ACTIVITY	Details ...	35	-0.24	-1.69	0.040	0.143
4	TRANSFERASE_ACTIVITY_TRANSFERRING_PHOSPHORUS_CONTAINING_GROUPS	Details ...	50	-0.18	-1.51	0.053	0.251

Figure 5.5: Enrichment of Molecular Functions GO terms for genes expressed differentially between E10.5 and E12.5.
(A) Number of differentially expressed genes for each comparison of developmental time points analysed. (B) and (C) Enrichment of Molecular Functions GO terms for genes significantly upregulated at (B) E12.5 compared to E10.5 and (C) E10.5 compared to E12.5. Terms were considered enriched if p-value <= 0.01 and q-value (FDR) <= 0.25. Enrichment plots and genes contributing towards the enrichment are shown for the Molecular Functions terms enriched at (B) E12.5 compared to E10.5.

furthest developmental time-points in the series, 1026 genes were up-regulated at E12.5, while 604 genes were down-regulated.

Foxn1 was not expressed significantly differentially between E11.5 and E12.5 (1.8 fold higher at E12.5, p-value 0.18), while its target genes *Dll4* (6.52 fold), *Ccl25* (8.35 fold), and *Cxcl12* (2.5 fold) showed significantly higher expression at E12.5 compared to E11.5 with p-values less than 0.01. This may be due to the large variation in expression of *Foxn1* observed between biological replicates (RPKM values for E11.5: 1.39 (rep1), 0.9 (rep2), and 2.56 (rep3); RPKM values for E12.5: 1.84 (rep1), 4.26 (rep2), 3.73 (rep3)), suggesting that more precise staging of embryos may be required to detect stage-specific changes in *Foxn1* expression. In this study, the embryos were staged based on days post coitum and a more rigorous staging using number of somites is recommended for the future. On the other hand, *Foxn1* expression was 8.75 fold higher at E11.5 (p-value = 6.25E-06) and 15.87 fold higher at E12.5 (p-value = 8.57E-09) than at E10.5.

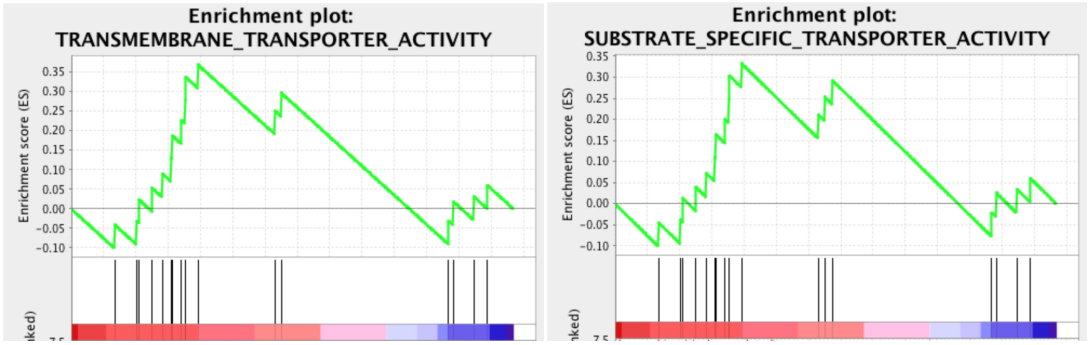
Studies on RNA-seq have not yielded a consensus on a method to determine the presence or absence of transcripts for genes within the samples being analysed. A common parameter used for this purpose is RPKM count for genes; however, different studies have used and recommended different thresholds for RPKM counts, ranging from 0.2 to 3 RPKM (Sultan et al. 2008; Ramsköld et al. 2009). These studies also showed that the optimal RPKM threshold varies between experiments (and between samples within an experiment) and various models have been proposed for determining experiment specific cutoffs (Ramsköld et al. 2009; Rau et al. 2013). For the purpose of this thesis, I have used a RPKM cut-off of 0.3, which has been proposed by Ramsköld and colleagues as the optimal cut-off which balances the numbers of false positive and false negatives across most samples used in their study (Ramsköld et al. 2009). Using this cut-off, *Gcm2* was expressed in the E10.5 (geometric mean of RPKM = 0.8) and E11.5 (geometric mean of RPKM = 4.2) samples but not in the E12.5 samples (geometric mean of RPKM = 0.17); whereas, *Pth* was expressed in the E11.5 (geometric mean of RPKM = 22) and E12.5

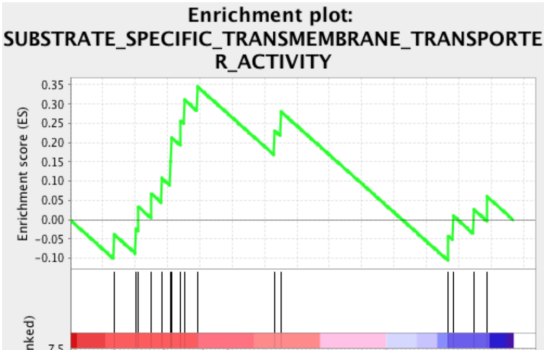
(geometric mean of RPKM = 1) samples but not in the E10.5 sample (geometric mean of RPKM = 0). This suggested that all of these samples contained some parathyroid cells, with the E11.5 samples showing a higher proportion of parathyroid cells than the other samples. Thus, some of the transcriptional changes described below might reflect this contamination. Using the threshold of 0.3 RPKM, 62.3% of the genes identified in Chapter 4 as having active promoters at E12.5 were designated as being expressed in E12.5 TEPCs, supporting the Chapter 4 analysis. On the other hand, 41% of the genes identified in Chapter 4 as being associated with putative active enhancers were expressed in E12.5 TEPCs, suggesting that identification of enhancer regions might suffer from a higher false positive rate.

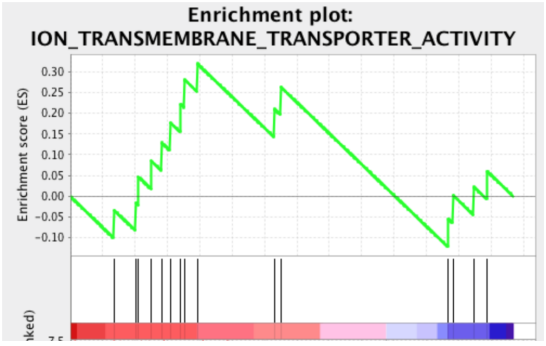
5.2.4.1 Enrichment analysis of differentially expressed genes

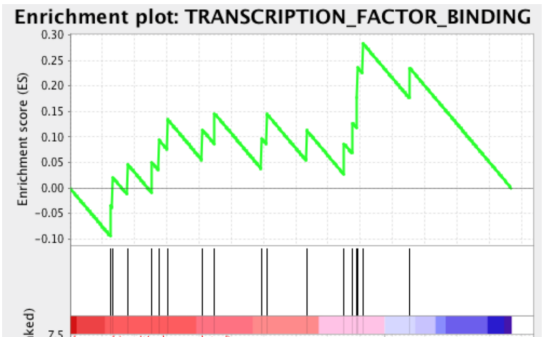
The Gene Set Enrichment Analysis (GSEA) software allows interpretation of gene expression data by performing enrichment analysis using gene sets available from the MSigDB database (Subramanian et al. 2005; Mootha et al. 2003; Liberzon et al. 2011). One of the key advantages of GSEA over similar methods is that it assesses the statistical significance associated with an observed enrichment by permuting the class labels, thus preserving gene-gene correlation important for biological processes and pathways. Secondly, the MSigDB database presents a comprehensive database of curated gene sets. On the other hand, a key limitation of enrichment analysis tools such as GSEA is that the ranking statistics used by such tools were selected for their effectiveness with DNA microarray data and their appropriateness for use with RNA-seq data has not been formally tested. Thus, instead of relying on GSEA's intrinsic ranking statisitcs, it is recommended to generate a pre-ranked list of genes from RNA-seq data, based on one of the test statistics generated from differential analysis, for use with GSEA. Analysis of molecular function enriched at E12.5 vs E10.5 comparison showed that genes up-regulated in E12.5 TEPCs were enriched (p-value < 0.05) for the GO terms "DNA Binding" and "transcription factor activity" at the 25% FDR (FDR cut-off as recommended by software developers) (Figure 5.5B). This suggests that E12.5 TEPCs are transcriptionally very different from E10.5 3PP cells, and that the expression of a large number of transcription factors is upregulated

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val
1	TRANSMEMBRANE_TRANSPORTER_ACTIVITY	Details ...	17	0.37	1.79	0.016	0.206
2	SUBSTRATE_SPECIFIC_TRANSPORTER_ACTIVITY	Details ...	18	0.33	1.69	0.034	0.163
3	SUBSTRATE_SPECIFIC_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	Details ...	16	0.35	1.67	0.034	0.127
4	ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	Details ...	15	0.32	1.51	0.070	0.200
5	TRANSCRIPTION_FACTOR_BINDING	Details ...	17	0.28	1.42	0.089	0.226
6	TRANSCRIPTION_FACTOR_ACTIVITY	Details ...	22	0.20	1.13	0.289	0.542









Genes contributing to enrichment of Transcription Factor Binding:
Nfia, Mafb, Kcnip3, Fosb, Ppargc1a, Junb, Atf3, Nab2, Srf, Mnt, Foxc1, Calcoco1, Lmo4, Nr2c2, Ddit3, Nfkbia

Figure 5.6: Enrichment of Molecular Functions GO terms for genes expressed differentially between E10.5 and E11.5.
Enrichment of Molecular Functions GO terms for genes significantly upregulated at E11.5 compared to E10.5. Terms were considered enriched if p-value <= 0.01 and q-value (FDR) <= 0.25. No enrichemnt was observed for genes upregulated at E10.5. Enrichment plots are shown for enriched terms. Also shown are the genes contributing towards the enrichment of "Transcription Factor Binding" term.

during development from E10.5 to E12.5. On the other hand, compared to E12.5 TEPCs, E10.5 3PP cells were enriched (p-value < 0.05) for genes annotated to the GO molecular function terms "Receptor Activity", "Transmembrane Receptor Activity" and "Protein Kinase Activity" at 25% FDR, suggesting significant changes in cell-surface receptors and signalling pathways between these two cell developmental stages (Figure 5.5C). Similarly, differential analysis of E11.5 vs E10.5 comparison showed enrichment for genes with transporter activity molecular function (Figure 5.6). Interestingly, the genes up-regulated at E12.5 compared to E11.5 (i.e. E12.5 vs E11.5 comparison) were not enriched for transcription-associated (or any other molecular function) terms (results from DAVID and GSEA).

Given that *Foxn1* expression in the thymic rudiment is initiated around E11.25, the transcription factors that are significantly up-regulated at E11.5 compared to E10.5 could be important for *Foxn1* expression. Analysis of genes significantly up-regulated at E11.5 vs E10.5 comparison showed that fifty-one TFs were up-regulated at E11.5 (results from DAVID). Of these, eleven transcription factors were found to have physical binding sites in the *Foxn1* promoter and enhancers regions identified in Chapter 4 (Table 5.1). These transcription factors are thus strong candidate regulators of fetal thymus development and of *Foxn1* expression in TEPCs. However, the expression of four out of the eleven TFs, namely *Gata3*, *Ebf1*, *Egr1*, and *Maf*, was significantly down-regulated at E12.5 compared to E11.5 (Table 5.1), suggesting that these TFs are not required for the maintenance of *Foxn1* expression and/or that their expression is repressed by FOXN1. These results for *Gata3* are consistent with the FOXN1-mediated repression of *Gata3* shown in Chapter 3. Furthermore, the expression of *Maf* and *Ebf1* were also up-regulated in E12.5 *nude* TEPCs compared to WT TEPCs (Stephanie Tetelin, Blackburn lab, unpublished), consistent with their repression by FOXN1.

Of the 51 TF genes up-regulated in E11.5 vs E10.5, only *Nfatc2* showed increased expression at E12.5 compared to E11.5 and had physical binding sites (NFAT1) in *Foxn1* promoter region identified in Chapter 4 (Table 5.1); although these were not

over-represented, further supporting it as potential transcriptional regulator of *Foxn1* in TEPCs. However, the expression profiles for *Nfatc2* (Figure 5.7) also suggest that this gene could be a target of *Foxn1*. This is consistent with the absence of expression of *Nfatc2* in E12.5 *nude* TEPCs (Stephanie Tetelin, Blackburn lab, unpublished). Thus, *Nfatc2* appears to be downstream of *Foxn1* in genetic network, rather than upstream. However, *Nfatc2* mutant mice models would be required to determine whether is involved in regulating *Foxn1* expression in TEPCs.

Of note, the RPKM values for *Nfatc2* were below the cut-off used above for determining whether a gene is expressed or not. However, the identification of differentially expressed genes using DESeq2 involves filtering out genes predicted to be not expressed, based on their raw read count (RC) values (Love et al. 2014). DESeq2 determines a dataset specific cut-off to determine the RC threshold and only returns Log2FoldChange (differential expression) for genes above this threshold (Love et al. 2014). Thus, it is likely that the RPKM cut-off of 0.3 could be too high for the dataset in this study, and that *Nfatc2* is expressed at low levels in the samples analysed here.

The gene *Nfatc2* is a member of the nuclear factor of activated T cells (NFAT) family. It encodes a DNA-binding protein with a REL-homology region (RHR) and an NFAT-homology region (NHR). The hyperphosphorylated form of NFATC2 protein is present in the cytosol and translocate to the nucleus upon dephosphorylation by phosphatase calcineurin. The NFAT family of proteins are important in many cellular processes, such as development and activation of lymphocytes, differentiation of cardiac muscle cells etc. (Crabtree & Olson 2002; Hogan et al. 2003). However, this transcription factor has not been previously implicated in TEC development.

Gene symbol	Gene	Binding site in Foxn1	diff expression E12.5 v/s E11.5
Ddit3	DNA-damage inducible transcript 3		
Ets1	E26 avian leukemia oncogene 1, 5' domain	yes	
Elf5	E74-like factor 5		
Fos	FBJ osteosarcoma oncogene		
Fosb	FBJ osteosarcoma oncogene B		
Gata3	GATA binding protein 3	yes	down
Isl1	ISL1 transcription factor, LIM/homeodomain	yes	
Jun	Jun oncogene		
Junb	Jun-B oncogene		up
Klf6	Kruppel-like factor 6		
Foxg1	forkhead box G1		
Tsc22d1	TSC22 domain family, member 1		
Atf3	activating transcription factor 3		
Arntl	aryl hydrocarbon receptor nuclear translocator-like		
Cux2	cut-like homeobox 2		
Dmrt2	doublesex and mab-3 related transcription factor 2		
Ebf1	early B-cell factor 1	yes	down
Egr1	early growth response 1	yes	down
Egr2	early growth response 2		
Esr1	estrogen receptor 1 (alpha)		
Esrrg	estrogen-related receptor gamma		down
Etv3	ets variant gene 3		
Foxc1	forkhead box C1	yes	
Foxn1	forkhead box N1		
Foxo1	forkhead box O1		
Fosl1	fos-like antigen 1		
Grhl1	grainyhead-like 1 (Drosophila); similar to Grhl1 protein		
Hoxa7	homeo box A7		
Hif3a	hypoxia inducible factor 3, alpha subunit		
Irf9	interferon regulatory factor 9		
Mnt	max binding protein		
Mef2c	myocyte enhancer factor 2C		
Nfia	nuclear factor I/A	yes	
Nfib	nuclear factor I/B	yes	
Nfatc1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1		
Nfatc2	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2	yes	up
Nr2c2	nuclear receptor subfamily 2, group C, member 2		
Nr4a1	nuclear receptor subfamily 4, group A, member 1		
Nr4a2	nuclear receptor subfamily 4, group A, member 2		down
Pax1	paired box gene 1		
Pax9	paired box gene 9		
Cphx	predicted gene 2135; predicted gene 2104; cytoplasmic polyadenylated homeobox		down
Rel	reticuloendotheliosis oncogene		
Srf	serum response factor	yes	
Maf	similar to c-Maf long form; avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog	yes	down
Mafg	similar to mafG; v-maf musculoaponeurotic fibrosarcoma oncogene family, protein G (avian)		
Six1	sine oculis-related homeobox 1 homolog (Drosophila)		
Ss18l1	synovial sarcoma translocation gene on chromosome 18-like 1		
Twist1	twist homolog 1 (Drosophila)		
Mafb	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian)		down
Zbtb38	zinc finger and BTB domain containing 38		

Table 5.1: Summary of transcription factors significantly upregulated at E11.5 compared to E10.5
Shown are all the transcription factors whose expression is significantly upregulated in E11.5 cells compared to E10.5 cells. Also indicated is whether these transcription factors have binding sites within the *Foxn1* promoter or enhancers identified in Chapter 4 and whether they are differentially expressed between E11.5 and E12.5.

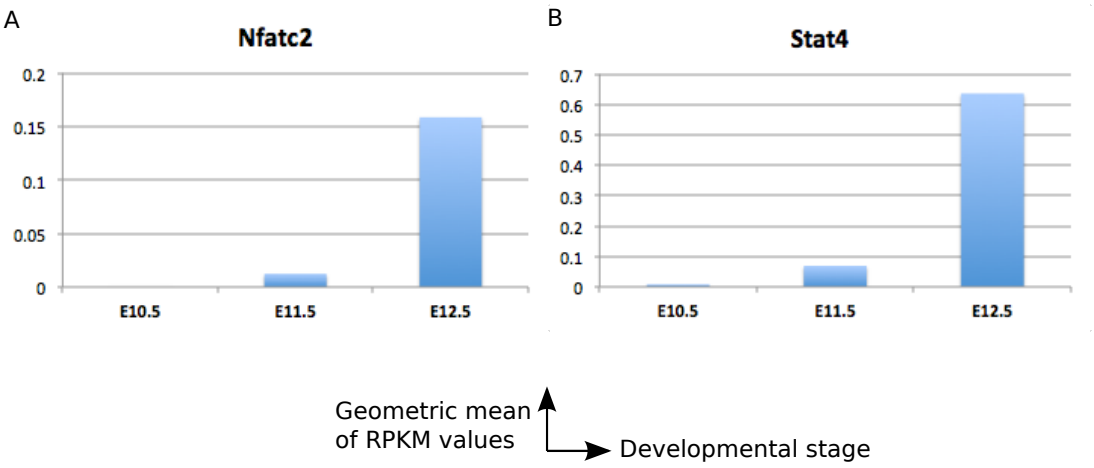


Figure 5.7: Expression of *Nfatc2* and *Stat4* in E10.5 and E11.5 third pharyngeal pouch endoderm cells and E12.5 thymic epithelial prgoenitor cells.
Shown are geometric mean of RPKM values for all biological replicates at each developmental time point analysed for *Nfatc2* (A) and *Stat4* (B).

5.2.4.2 Identification of co-regulation of differentially expressed genes

Genes involved in a genetic network are often co-regulated by one or more transcription factors. Such transcription factors regulating the expression of a set of genes can be identified by enrichment analysis of TFBSs within the promoter regions of a set of genes. In order to determine whether the genes that are differentially expressed between E11.5 and E10.5 might be co-regulated, I analysed the enrichment of TFBSs present in the promoter regions of these genes using GSEA. This analysis revealed a significant enrichment in genes that are up-regulated at E11.5 for binding motifs corresponding to NFATC2 (p – value = 0.002; q-value = 0.248) and STAT4 (p-value = 0.008; q-value = 0.117) at a FDR ≤ 25%. As mentioned above, the *Nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 2* (*Nfatc2*) gene, which encodes the NFATC2 protein, is itself significantly up-regulated at E11.5 and also has putative binding sites within the *Foxn1* promoter (Table 5.1). A closer analysis of the above results showed that sixty-six genes significantly up-regulated at E11.5 compared to E10.5 contributed to the enrichment of NFATC2 motif (Table 5.2), suggesting that the up-regulation of these genes could result from increased NFATC2 activity. Thus, this transcription factor is likely to be important for thymus development.

In the above analysis, fifteen genes contributed to the enrichment of STAT4 motif (Table 5.2). The TFBS analysis in Chapter 4 showed that the identified *Foxn1* promoter and enhancers did not contain any STAT4 binding sites. The Janus Kinase-Signal Transducers and Activators of Transcription (JAK-STAT) pathways play critical roles in the immune, neuronal, hematopoietic and hepatic systems. The STAT family of transcription factors are phosphorylated by the receptor-associated kinases in response to cytokines and growth factors. Phosphorylated STAT proteins then form homo- or heterodimers that translocate to the cell nucleus to regulate their target genes. *Stat4* was significantly up-regulated in E12.5 TEPCs compared to both E10.5 and E11.5 3PP cells and was also up-regulated at E11.5 compared to E10.5, although this up-regulation was not statistically significant (p-value = 0.015; padj

(FDR) = 0.21). The expression profile for *Stat4* is shown in Figure 5.7. Similar to *Nfatc2*, the expression of *Stat4* was absent in E12.5 *nude* TEPCs (Stephanie Tetelin, Blackburn lab, unpublished), suggesting *Stat4* as a target of *Foxn1*.

Finally, analysis of genes expressed differentially between E12.5 and E11.5 or E10.5 showed that genes down-regulated at E12.5 were enriched for binding motifs for a number of different transcription factors. These TFs include: LEF1, E12, FOXO4, FOXO1, MAZ, FOXC1, FOXF2, NFAT1, TST1, MYOD, STAT5A (for E12.5 vs E11.5); and LEF1, HNF1, AP3, TAL1, E12, CEBPGAMMA, SOX9, ZID, FOXD1, FOXO4 (for E12.5 vs E10.5). Given the enrichment of the NFAT1 motif in the genes down-regulated at E12.5 compared to E11.5, it was important to determine whether there was an overlap between these genes and the genes contributing to NFAT1 motif enrichment in E11.5 vs E10.5 analysis described above, in order to determine whether genes with an NFAT1 motif are transiently up-regulated at E11.5. There were sixty-four genes significantly down-regulated at E12.5 compared to E11.5 that contributed to the enrichment for the NFAT1 motif. Of these, only eight genes were shared between the two lists: *Esr1*, *Il1rapl1*, *Spag6*, *Flrt3*, *Fst*, *Gata3*, *Esrrg*, and *Khdrbs2*. The absence of a significant overlap between these two sets of genes with NFAT1 motif suggested that NFAT1 may be involved in activating or suppressing the expression of different sets of genes during thymus development.

The enrichment of LEF1 and E12/TCF3 motifs in genes down-regulated at E12.5 compared to both E11.5 and E10.5 suggests that there is a dynamic regulation of the Wnt signalling pathway during these stages. Among the Wnt ligands, *Wnt4* was expressed in E12.5 TEPCs (geometric mean of RPKM = 2) and in E11.5 3PP cells (geometric mean of RPKM = 0.7), but not in E10.5 3PP cells (geometric mean of RPKM = 0.1). No other Wnt ligands were expressed at any of the time points analysed. The expression of *Fzd1* decreased from E10.5 (geometric mean of RPKM = 12) to E12.5 (geometric mean of RPKM = 5), with a spike at E11.5 (geometric mean of RPKM = 23). On the other hand, *Fzd3* was not expressed at any stage and the expression of *Fzd4* was detectable at E10.5 (geometric mean of RPKM = 2) but

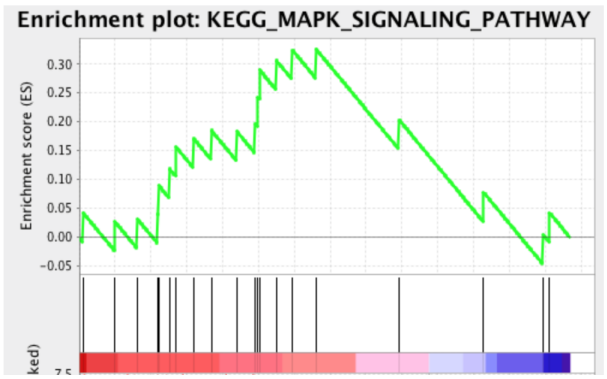
not at E11.5 or E12.5. Finally, the expression of *Gsk3b* increased from E10.5 (geometric mean of RPKM = 4.5) to E11.5 (geometric mean of RPKM = 9.7) and remained unchanged between E11.5 and E12.5 (geometric mean of RPKM = 8.65). The increase in *Gsk3b* expression after E10.5 suggests that the activity of Wnt signalling pathway is downregulated after E10.5. Thus, E11.5 3PP cells and E12.5 TEPCs appear to be less able to respond to Wnt signalling compared to E10.5 3PP cells.

The above analyses of differentially expressed genes revealed important insights into transcriptional changes associated with development of E10.5 3PP cells to E12.5 TEPCs. The TFs whose expression is significantly up-regulated at E11.5 compared to E10.5, hence strong candidate regulators of *Foxn1*, were identified. The above results suggest an important role of *Nfatc2* in thymus development, evidenced by the enrichment of its binding motif in the promoters of genes regulated differentially between the analysed time points. An important limitation of the above analyses is the quality of PWMs for the TFs being analysed. This is likely to affect the enrichment for TFs such as FOXN1 and PAX1, which lack a good quality PWM but are known to be important for thymus development.

5.2.5 Identification of differentially regulated signalling pathways

To determine additional functional and mechanistic differences in 3PP cells and TEPCs isolated from various developmental time points, the list of differentially expressed genes obtained from DESeq2 was used as input data for Generally Applicable Gene-set Enrichment analysis (GAGE) (Luo et al. 2009) and Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005; Mootha et al. 2003). Pathway enrichment was carried out against Kegg pathways using these two software tools. An advantage of the GAGE package is its integration with the RNA-seq data analysis pipeline, as it can take the *results* object generated using DESeq function as input data. Furthermore, unlike other methods, GAGE allows identification of gene sets whose genes show differential expression in both directions (i.e. gene could either be up- or down-regulated), which is frequently the

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val
1	KEGG_MAPK_SIGNALING_PATHWAY	Details ...	20	0.33	1.73	0.024	0.026



Genes contributing to enrichment of MAPK signalling:
Nr4a1, Mef2c, Fos, Dusp1, Nfatc2, Dusp14, Dusp2, Fgf12, Dusp10, Il1r1, Map3k5, Dusp5, Map3k14, Jun, Srf, Mapk8ip1

Figure 5.8: Enrichment of signalling pathway at E11.5 compared to E10.5
Analysis of signalling pathway enrichment showed enrichmnet of MAPK signalling at E11.5 compared to E10.5. Terms were considered enriched if p-value <= 0.01 and q-value (FDR) <= 0.25. Shown are p-value, q-value, enrichment plot, and genes contributing towards enrichment of this pathway.

case for canonical signalling pathways. On the other hand, GSEA has been used extensively for analysing gene set enrichments from gene expression and its "Preranked workflow" allows its use in RNA-seq data analysis. This workflow takes a list of genes and associated numerical values (such as fold change, p-value, etc.) as an input and sorts the genes based on the values in the numerical column for enrichment analysis.

GSEA or GAGE analysis of genes differentially expressed at E12.5 vs E11.5 did not identify any signalling pathways enriched at FDR ≤10% for GAGE and FDR ≤25% for GSEA. Similarly, GAGE analysis of genes expressed differentially between E11.5 and E10.5 did not identify any significantly enriched pathway at FDR ≤10%. On the other hand, the GSEA analysis of these genes showed that they are enriched for genes involved in MAPK signalling pathway (p-value = 0.024; q-value = 0.026). Figure 5.8 shows the enrichment plot and the differentially expressed genes involved in this pathway. There appears to be an increase in the activity of MAPK signalling at E11.5 compared to E10.5.

GSEA identified three pathways that were enriched at E10.5 versus E12.5 at a 25% FDR: the "Axon Guidance" (p-value = 0.01; q-value = 0.053), "Focal Adhesion" (p-value = 0.013; q-value = 0.069), and "Regulation of Actin Cytoskeleton" (p-value = 0.046; q-value = 0.112) pathways. The enrichment values and plots for these pathways are shown in Figure 5.9. The importance of these pathways in thymus development is yet to be determined.

On the other hand, GAGE analysis of the E12.5 vs E10.5 data showed significant enrichment for "mmu04064 NF-Kappa B Signalling Pathway" (p-value=0.0003 and q-value=0.071) at a FDR ≤10%. The enrichment of NFκB signalling in E12.5 TEPCs is consistent with the enrichment for this pathway observed in Chapter 4 (section 4.6.1). Enriched pathways identified using the GAGE package can be visualized using a pathway data visualization package called *pathview* (Luo & Brouwer 2013). Pathview allows integration of user data onto the relevant pathway and supports

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val
1	KEGG_FOCAL_ADHESION	Details ...	25	-0.33	-1.90	0.010	0.053
2	KEGG_AXON_GUIDANCE	Details ...	27	-0.29	-1.73	0.013	0.069
3	KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	Details ...	25	-0.26	-1.56	0.046	0.112
4	KEGG_WNT_SIGNALING_PATHWAY	Details ...	20	-0.23	-1.21	0.207	0.346

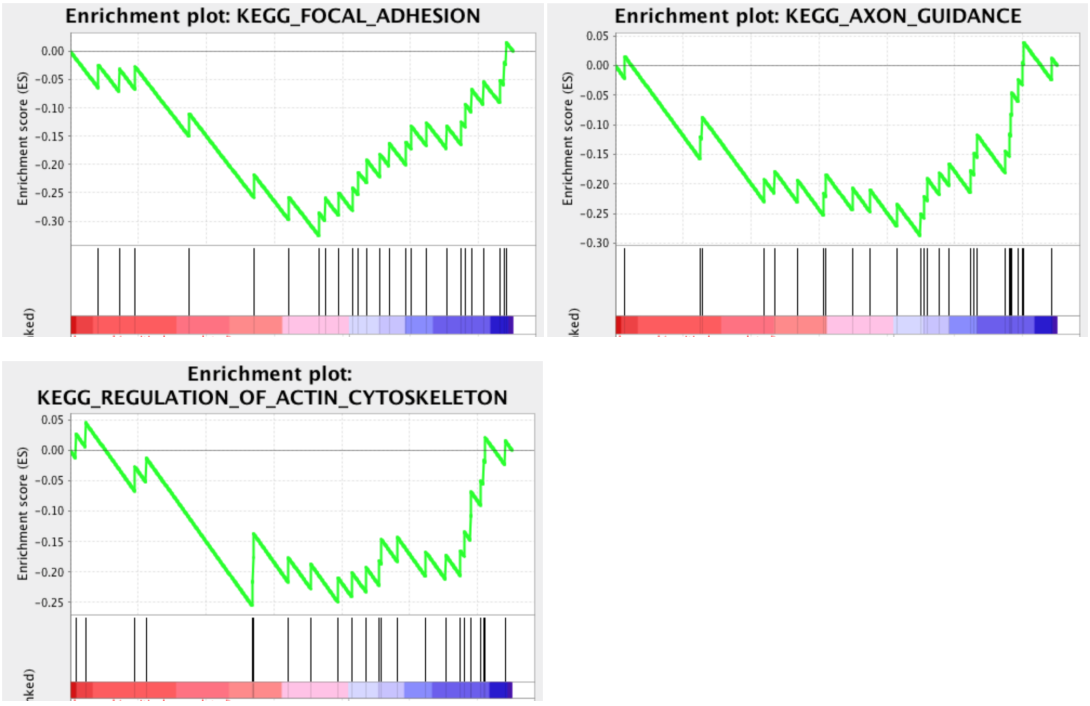


Figure 5.9: Enrichment of signalling pathways at E10.5 compared to E12.5
Analysis of signalling pathway enrichment showed enrichment of Focal Adhesion, Axon Guidance, and Regulatin of Actin Cytoskeleton at E10.5 compared to E12.5. Terms were considered enriched if p-value <= 0.01 and q-value (FDR) <= 0.25. Shown are p-values, q-values and enrichment plots for enriched pathways.

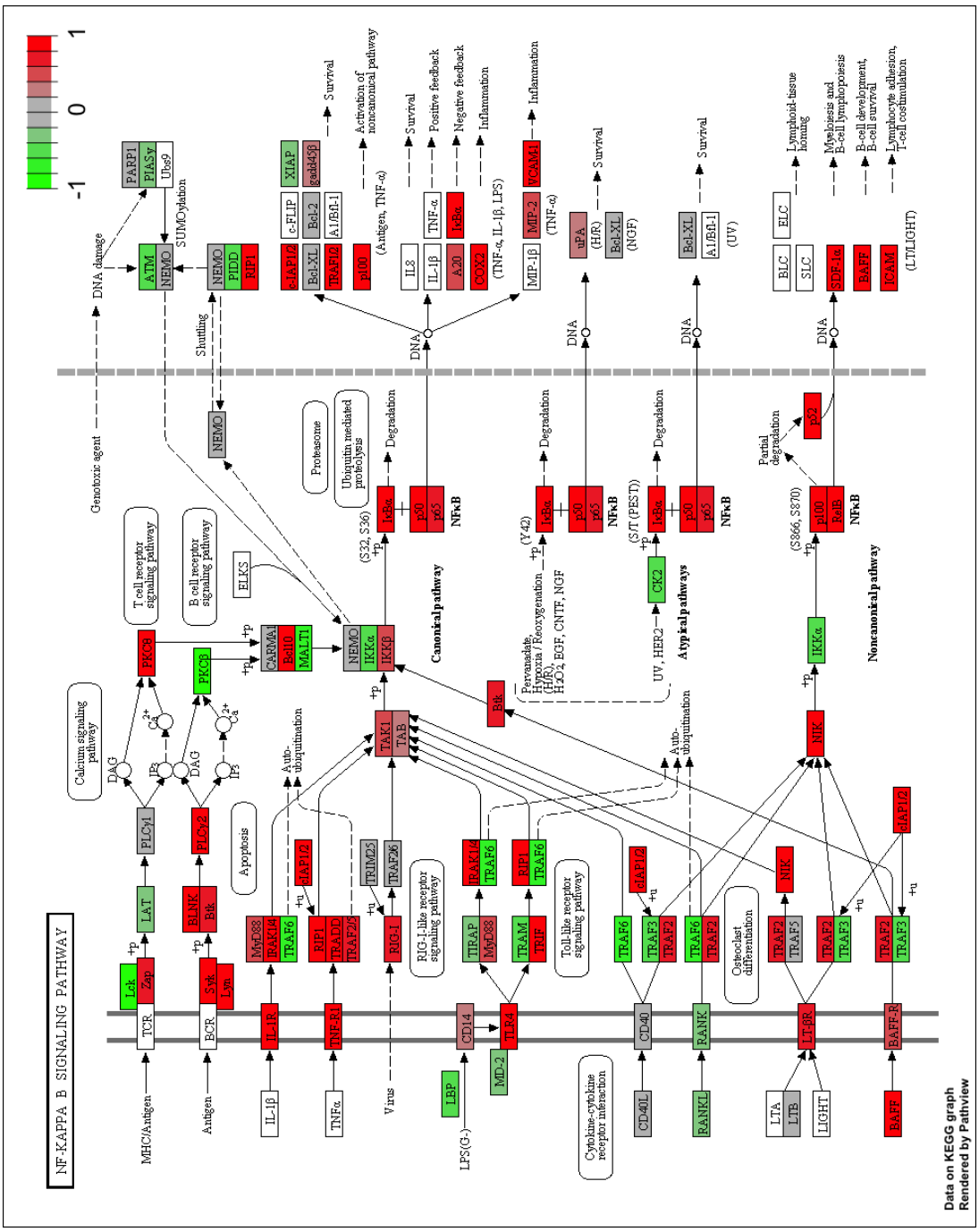


Figure 5.10: Kegg view of NF-kB pathway enrichment at E12.5 compared to E10.5
GAGE algorithm was used to analyse pathway enrichment between E12.5 and E10.5. GAGE determines pathway enrichment using fold differences for all genes. Genes contributing towards the observed enriched are shown in the Kegg graph with either green, red, or grey colour. Genes with higher expression in E12.5 TEPCs are shown in red and those with higher expression in E10.5 are shown in green. The colour intensities correspond to differences in expression between E10.5 and E12.5.

visualization of the rendered pathway graphs with mapped data. Figure 5.10 show the differential gene expression data rendered onto the Kegg view of the NFκB signalling pathway. As evident from this graph, the expression of a number of genes involved in this pathway changes between E10.5 and E12.5. The genes highlighted in red are more highly expressed in E12.5 TEPCs, while those highlighted in green are expressed more highly in E10.5 3PP cells. Based on this, the pathway seemed to be more active in E12.5 TEPCs compared to E10.5 3PP cells.

The rendered Kegg pathways provide a visual overview of the changes in expression of all the genes involved in the pathway. However, it is important to note that the GAGE analysis performed here only utilizes log2FoldChange and does not account for p-value or FDR. Thus, while the colour codes (red and green) represent the general direction of difference in expression between two samples, they do not necessarily represent statistically significant differences in gene expression. An advantage of rendering the data onto Kegg pathways is that the resulting pathway view retains pathway meta-data such as protein interactions, spatial information for proteins within a cell, and interactions with other signalling pathways. This is evident from the various interactions between NFκB and other signalling pathways shown in Figure 5.10. This figure suggests that there is an up-regulation at E12.5 compared to E10.5 of genes involved in cell survival (such as *Birc3*, *Gadd45b*, *Traf1*, and *Traf2*), non-canonical NfκB pathways (such as *p100*/*Nfkb2*), feedback mechanisms (such as *A20/Tnfaip3* and *Nfkbia/I*κBα) and lymphocyte attraction and activation (such as *Cxcl12*/*Sdf1*, *Cxcl2*/*Mip2*, *Vcam1*, *Baff*, *Icam1*, and *Cox2/Ptgs2*). Of these, *Nfkb2* (log2FoldChange = 2.7), *Cxcl12* (log2FoldChange = 1.5), *Vcam1* (log2FoldChange = 3.87), *Icam1* (log2FoldChange = 2.88), *Nfkbia* (log2FoldChange = 2.98), and *Ptgs2* (log2FoldChange = 3.88) were significantly up-regulated at E12.5 compared to E10.5 at FDR ≤ 10%. Thus, activation of NFκB signalling in E12.5 TEPCs causes significant up-regulation of negative feedback gene (*Nfkbia*) and genes promoting the non-canonical pathway (*p100*), suggesting that similar to observations in other cell types, the canonical NFκB signalling pathway might oscillate between active and inactive state in developing TECs. The remaining genes too were up-regulated, however, their fold increase was not significant at FDR ≤ 10%. The log2FoldChange

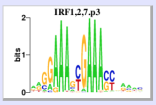
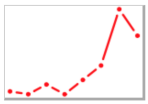
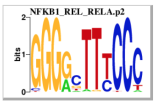
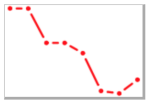
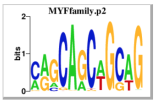
in the expression of these genes was as follows: *Birc3* (1.5), *Gadd45b* (0.44), *Traf1* (0.31), *Traf2* (0.75), *Cxcl2* (0.53), *Baff*/*Tnfsf13b* (1.45), *Tnfaip3*/*A20* (0.45).

A number of signalling pathways converging on NFκB activation have been shown to be important in mTEC development, particularly those involving the tumour necrosis factor family of receptors (Rossi et al. 2007; Akiyama et al. 2008; Hikosaka et al. 2008). A closer analysis of the receptor involved in activation of the NFκB pathway, as shown in Figure 5.10, showed significantly up-regulated expression (at FDR ≤ 10%) of the *Interleukin 1 receptor 1* (*Il1r1*; log2FoldChange = 4.67) and *Toll like receptor 4* (*Tlr4*; log2FoldChange = 2.53) genes at E12.5 compared to E10.5. However, the reduction (although statistically insignificant) in expression of *Lbp* and *Md2* (Figure 5.10) at E12.5 compared to E10.5, suggests that IL1R1 is the major activator of NFκB signalling in these cells. While activation of the NFκB signalling pathway in E12.5 TEPCs appears to be mediated through activation of one or both of these two receptors, its importance and function in TEPCs remains unknown. These pathway analyses suggest that there is increased activation of NFκB in E12.5 TEPCs as compared to E10.5 3PP cells and thus that this pathway is likely to be involved in subsequent thymus development. Of note is that analyses in Chapter 4 identified several binding sites for the p65/RELA protein, which is involved in NFκB signalling, within the *Foxn1* promoter and putative enhancers identified in that chapter. Thus, the role of NFκB signalling in thymus development warrants further investigation.

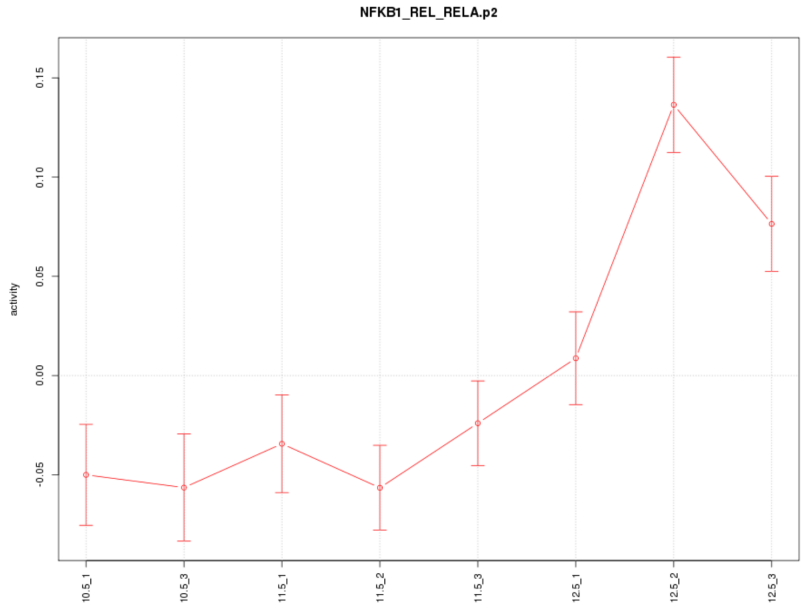
Figure 5.11: Analysis of transcription factor activities using RNA-seq.

See following page
(A) Top four transcription factor motifs with highest Z-score in ISMARA results. The Z-score is a representation of significance of a motif's activity. Higher score indicate more significance. Also shown are transcription factors associated with the respective motifs, the motif activity profile (left to right: E10.5 to E12.5), and the matrix logo representing the position weight matrix. (B) Line graph demonstrating motif activity, for motifs with increasing activity from E10.5 to E12.5, in the analysed samples. Order of samples on X-axis: (left to right) E10.5 rep1, E10.5 rep3, E11.5 rep1, E11.5 rep2, E11.5 rep3, E12.5 rep1, E12.5 rep2, and E12.5 rep3. Top graph shows motif activity for IRF1,2,7 motif and bottom graph shows activity for NFKB1_REL-RELA motif.

A

Motif name	Z-value	Associated genes	Profile	Logo
IRF1,2,7.p3	4.256	Irf1 (Irf-1) Irf2 (Irf-2) Irf7		
NFKB1_REL_RELA.p2	2.762	Rela (p65 NF kappaB, p65) Rel (c-Rel) Nfkb1 (nuclear factor kappaB p50, NF-kappaB p50, p50, NF-kappaB, p50/p105, p50 subunit of NF kappaB, NF kappaB1)		
HOX{A6,A7,B6,B7}.p2	2.419	Hoxb6 (Hox-2.2) Hoxa7 (Hox-1.1) Hoxb7 (Hox-2.3) Hoxa6 (Hox-1.2)		
MYFfamily.p2	2.158	Myod1 (Myod-1, bHLHc1, MyoD, MYF3) Myf6 (herculin, MRF4, bHLHc4) Myf5 (bHLHc2, Myf-5) Myog (bHLHc3, myo)		

B



5.2.5.1 Identification of transcription factor networks in TEPCs

To identify networks of transcription factors important in TEPCs, I further analysed the RNA-seq data using the Integrated System for Motif Activity Response Analysis (ISMARA) tool, which models gene expression in terms of genome-wide predictions of regulatory sites to make detailed predictions regarding the roles of key TFs and direct interactions between regulators (Balwierz et al. 2014). ISMARA takes mapped reads as an input and then quantifies the activity of promoters in each sample (defined in ISMARA as 1000bp region centred on TSS) by determining the logarithm of the estimated number of transcripts per million transcripts (similar to RPKM). Direct regulatory interaction between motifs is predicted between two motifs when one of the motifs is predicted to target the promoter of at least one of the TFs associated with the second motif (Balwierz et al. 2014). ISMARA provides these predictions as a local network picture for individual motifs that show all predicted regulatory connections with other motifs (Balwierz et al. 2014).

Analyses of the RNA-seq data generated in this chapter using ISMARA identified motifs for IRF and NFKB/REL TFs as the two most significant motifs (Figure 5.11A; Z-score represents the importance of the motif for explaining expression variation across the sample). This analysis predicted an increase in the activity of NFKB motif at E12.5 compared to E10.5 and E11.5 (Figure 5.11B), consistent with the enrichment of the NFκB signalling pathway described above. Similarly, the IRF motif activity was also predicted to increase from E10.5 to E12.5 (Figure 5.11B). ISMARA also determines correlation coefficients between the expression of the TFs associated with a motif and its predicted activity (activity is based on the changes in expression of a motif's predicted target genes). The expression of all three genes associated with the IRF motif, *Irf1*, *Irf2* and *Irf7*, showed significant correlation with its predicted activity (Figure 5.12A; Figure 5.12B-D). On the other hand, the genes *Rela* and *Nfkb1*, but not *Rel*, showed significant correlation between their expression and the activity of the NFKB motif (Figure 5.12A; Figure 5.12E-F).

Finally, ISMARA provides a graphical view of predicted direct regulatory interactions between motifs, i.e. transcription factors network. Thus, I investigated the regulatory interactions of significant motifs (score threshold of 0.9) to identify TF networks involving the genes associated with these motifs. The predicted regulatory interaction for NFKB motif suggested *Irf1* and *Nfatc2* as targets of TFs associated with NFKB motif (Figure 5.13A). Such interaction between NFκB signalling and *Irf1* gene has been demonstrated previously in other cell type (Harada et al. 1994). Furthermore, the transcription of *Nfatc2* in TEPCs appears to be under the control of NFκB signalling. Similarly, the predicted regulatory interactions for IRF motif suggested *Arid5b*, *Stat1*, and *Nfkb1* as direct targets of the TFs associated with this motif (Figure 5.13B). Furthermore, both NFKB and IRF motifs showed positive feedback on the expression of associated TFs: *Nfkb1* and *Rel* in the case of NFKB motif and *Irf7* in case of IRF motif.

Another significant motif identified from the above analysis was a HOX motif (Figure 5.11A). However, the activity of the Hox motif showed an opposite profile compare to that for the IRF and NFKB motifs, i.e. the predicted activity of HOX motif decreased from E10.5 to E12.5. Among the *Hox* genes associated this motif, the expression of only *Hoxa7* showed a significant (anti)correlation with the motif activity (i.e. expression of its predicated target genes) (Correlation Coefficient: -0.74; p-value = 3.4E-02). These results suggested that the expression of HOX motif target genes decreased from E10.5 to E12.5 and that this decrease was at least partly a result of increased expression of *Hoxa7*, which acts a repressor for the HOX motif target genes. The predicted network interaction for the HOX motif suggested that the TFs associated with this motif could act as transcriptional regulators of *Nr2f2*, *Tcfap2*, *Sox2*, *Sox5*, *Msx2*, *Hoxb5*, *Hoxa5*, *Nhlh2*, *Tcfap2b*, and *Prrx1* (Figure 5.13C).

The above results therefore suggest that the development of the thymic primordium from E10.5 to E12.5 is accompanied by a reduction in the activity of a TF network regulated by HOXA7 and an increase in activity of TF networks regulated by Irf and

NFKB. The ISMARA analysis did not predict an interaction between the IRF/NFKB and HOX motifs, thus the prevalence of potential crosstalk between the TF networks involving these motifs remains to be determined. It is important to note that these predicted regulatory interaction networks were generated using an arbitrary threshold for significance values associated with individual interactions and changing this threshold leads to increase or decrease in the number of predicted interactions.

5.3 Discussion

The RNA-seq data described in this chapter provide an opportunity to study transcriptional changes associated with developing and differentiating TEPCs. Indeed, comparison of the transcriptomes of 3PP cells and TEPCs isolated from E10.5, E11.5, and E12.5 identified thousands of differentially expressed genes. An important caveat in the present study is the presence of parathyroid cells in the collected samples. The thymus and parathyroid are generated from a common primordium arising from 3PP and only begin to separate around E12.0. This common primordium expresses *Plet1* (Depreter et al. 2007), which was used for sorting cells for RNA-seq. Thus, the sorted cells also contain some proportion of parathyroid cells, although this proportion is likely to be small as the thymus fated domain of 3PP is larger than the parathyroid fated domain. However, it is not possible to confidently determine the contribution of parathyroid cells to the gene expression

Figure 5.12: Analysis of contribution of transcription factors to motif activity.

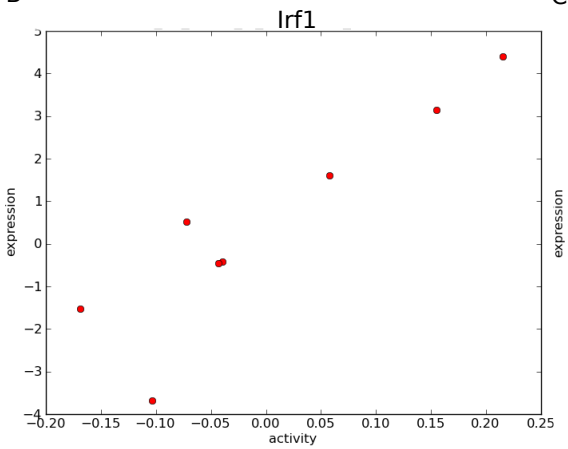
See following page

(A) Pearson correlation coefficient representing the correlation between motif activity and the gene expression of associated transcription factors. Positive correlation coefficient indicates an increase in gene expression for a transcription factor with an increase in motif activity. The p-values represent the significance of the observed correlation. (B), (C), (D), (E), and (F) Correlation graphs showing transcription factor gene expression on Y-axis and motif activity in X-axis for (B) *Irf1*, (C) *Irf2*, (D) *Irf7*, (E) *Nfkb1*, and (F) *Rela*. Graph for Rel is not shown as its expression is not significantly correlated to motif activity.

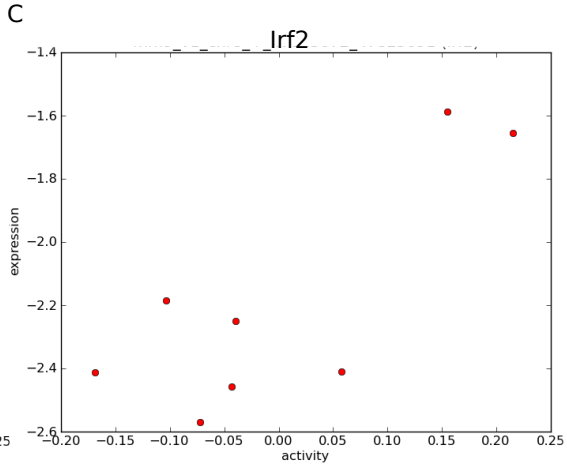
A: coorelation of motif activity with associated transcription factor gene expression

Gene Symbol	Pearson corr. coef.	p-value
IRF motif		
Irf1	0.91	1.80E-03
Irf7	0.83	1.00E-02
Irf2	0.8	1.70E-02
NFKB motif		
Rela	0.84	8.80E-03
Nfkb1	0.84	9.20E-03
Rel	0.5	2.00E-01

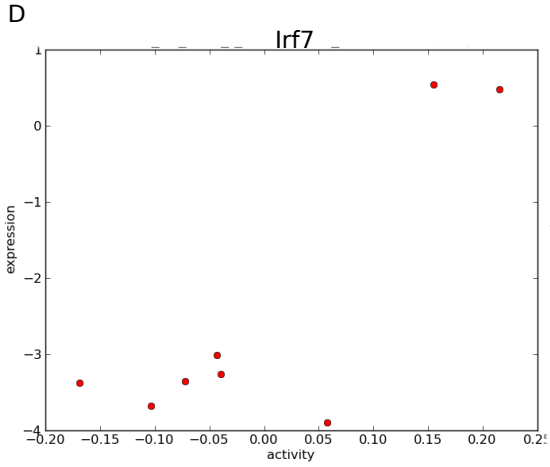
B



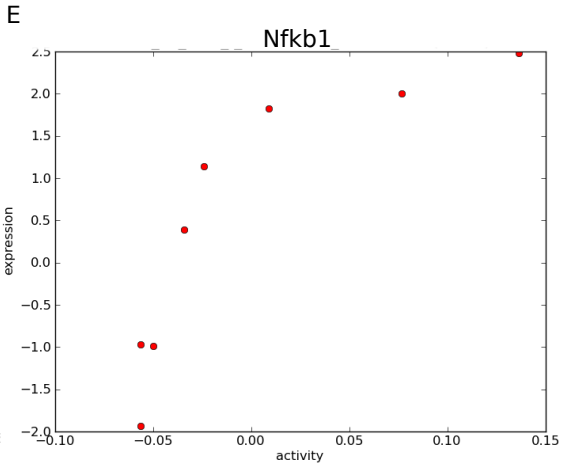
C



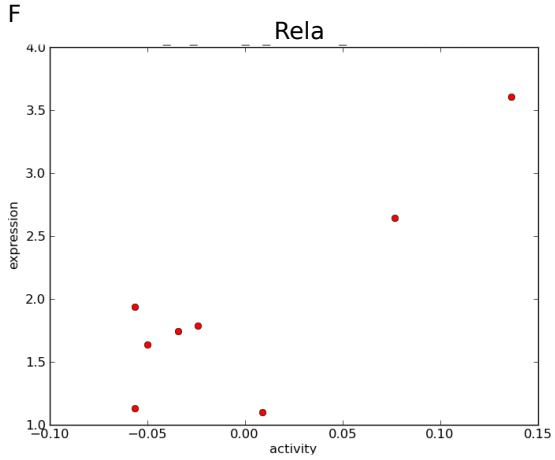
D



E



F



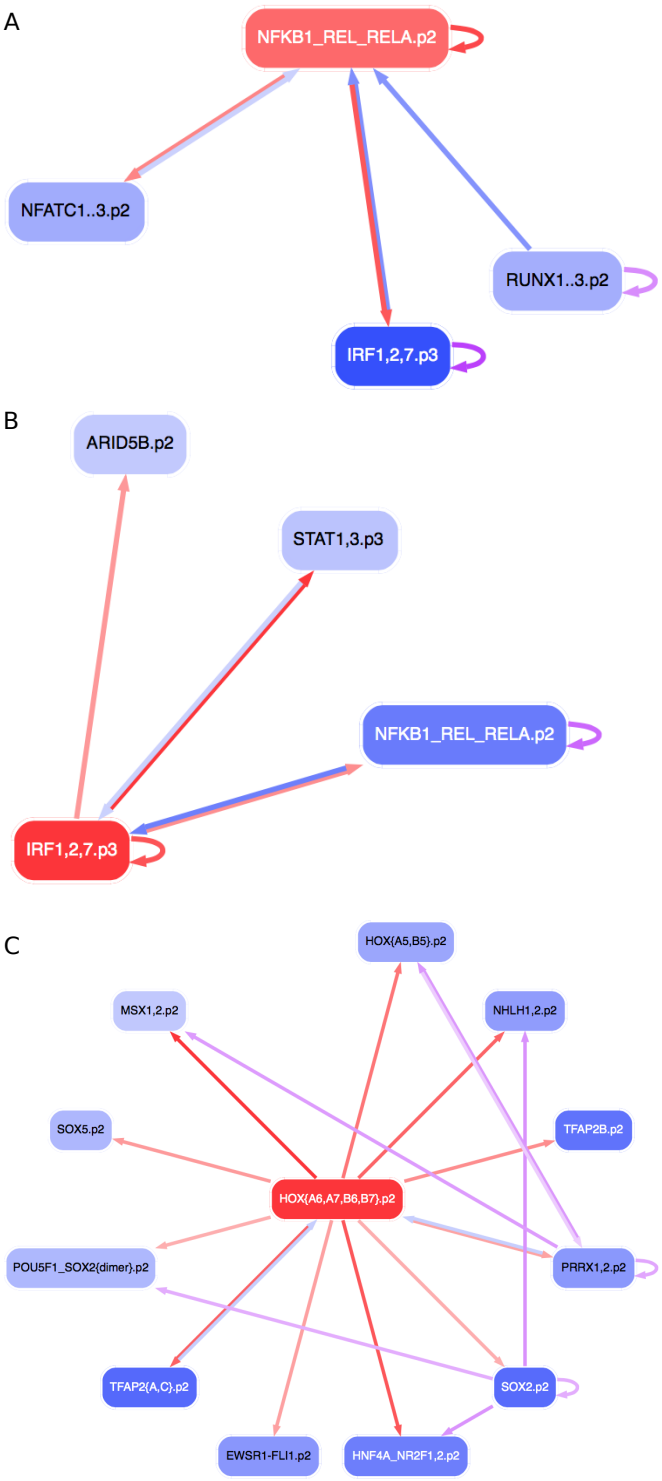


Figure 5.13: Predicting transcription factor network from RNA-seq.
Transcription factors network diagram showing predicted interactions between transcription factors. Interactions are predicted based on the presence of binding sites for transcription factor associated with one motif within the promoters of genes associated with another motif. Shown here are networks centered on NFKB1_REL_RELA (A), IRF1,2,7 (B), and HOX{A6,A7,B6,B7} (C) motifs. Only interactions with a Z-score of at least 0.9 are shown here. The colour intensities are representative of Z-score, higher intensity representing higher Z-score. Regulation by central motif are shown in shades of red, while regulation of central motif by other motifs is shown in blue. Interactions between motifs not involving central motifs are also shown.

changes observed here. Interestingly, no clear relationship between genes that are in the vicinity of an active enhancer region and genes showing a change in expression during development was observed upon comparison of these two sets. This could be due to a less than optimal filtering of genes with no changes in expression, this set contains more than 15,000 genes, and/or a high false discovery rate for active enhancers identified in Chapter 4.

The results from this Chapter are summarized in Figure 5.14. The results shown in Figure 5.4 show that clustering of genes sharing similar expression profiles distinguishes between the three different developmental stages analysed, E10.5, E11.5 and E12.5. The enrichment of genes associated with thymocyte development at E12.5 suggests that even though *Foxn1* expression is initiated at E11.25, the TEC differentiation programme is only initiated at around E12.5. This is consistent with the developmental arrest of *nude* thymic rudiment at E12.5 and with data from Nowell and colleagues (Nowell et al. 2011). This result also suggests that the genes present in Cluster 1 (and other genes with similar expression pattern) could be important for TEPC differentiation and might represent FOXN1 target genes or genes that act together with FOXN1 to promote TEPC differentiation. The genes in Cluster 2 show expression profiles similar to that of *Foxn1*. This profile is consistent with that expected of *Foxn1* transcriptional regulators (in the absence of feedback loops) or early FOXN1 targets. Cluster 3 represents genes that are transiently up-regulated at E11.5. A pulse of BMP signalling has been shown to be important for initiation of *Foxn1* expression (Neves et al. 2012), whether a similar requirement exists for Cluster 3 genes remains to be tested. It is important to note, however, that the E11.5 samples showed the highest expression of *Pth* (and *Gcm2*) and this gene is indeed present in Cluster 3. Thus it is possible that some of the genes in Cluster 3 are required for parathyroid development and shows a subsequent down-regulation at E12.5 due to presence of fewer parathyroid cells in these samples. Finally, the genes in Cluster 4 are enriched for cell fate specification and organ development terms, among others, and are thus likely to represent genes important for formation of 3PP and its subsequent commitment to thymus and parathyroid fates. However, further studies are required to understand the roles of these genes. There is also a possibility

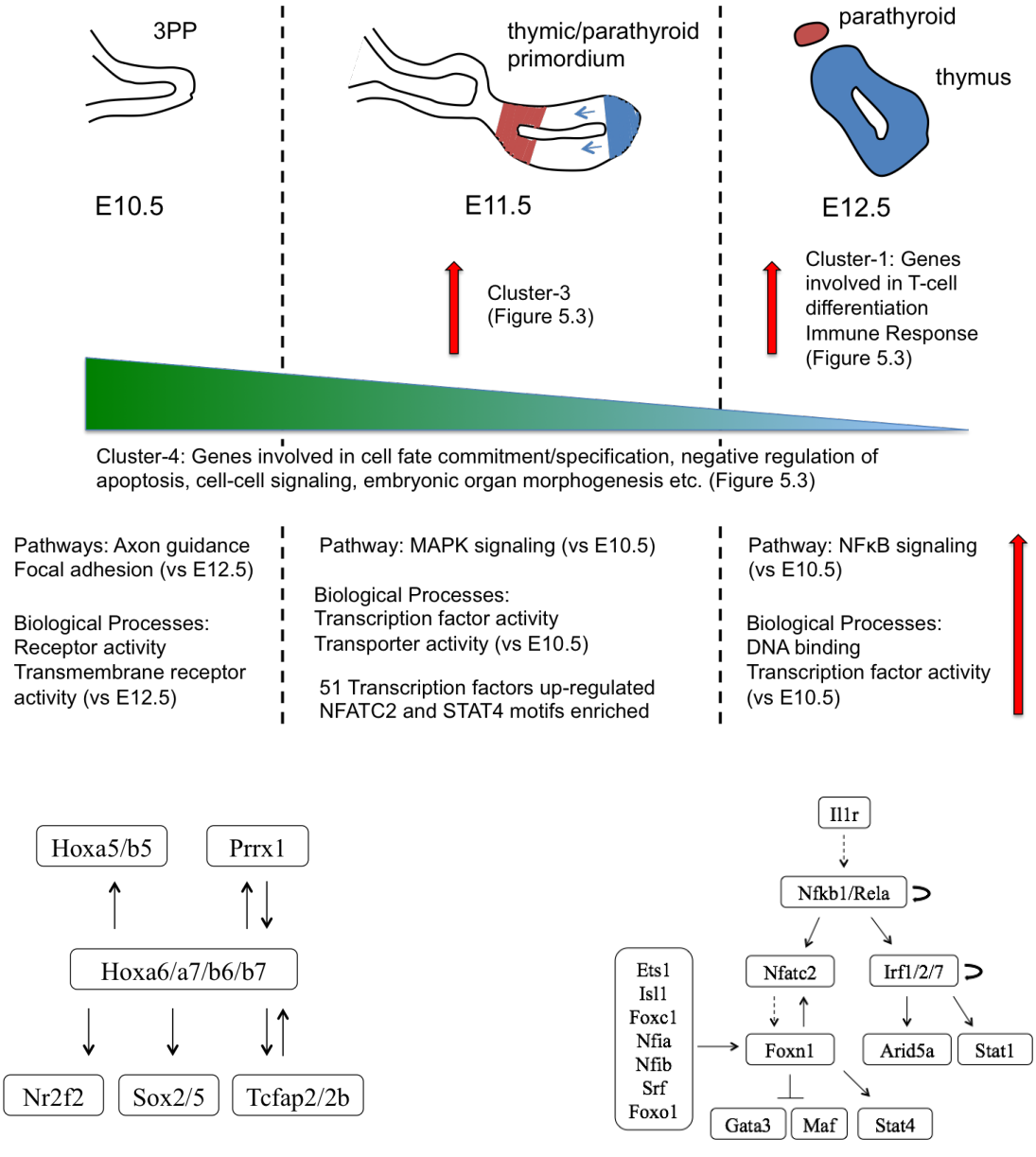


Figure 5.14: Summary of Chapter 5 results.
Summary of all the results obtained from comparison of E10.5 and E11.5 3PP cells and E12.5 TEPCs transcriptomes. Shown are the changes observed in gene expression, molecular functions, biological processes, signalling pathways, and transcription factor network during thymus developement.

of these genes acting as transcriptional repressors of genes involved in Cluster 2 (or vice versa), although this hypothesis can not be tested using the present data.

The expression of *Msx2*, shown to be required for normal *Foxn1* expression in hair follicle (Cai et al. 2009), was down-regulated from E10.5 to E12.5, suggesting that it is not required for *Foxn1* expression in the thymus. Analysis of differentially expressed genes revealed several interesting patterns in the data. Firstly, compared to E10.5, both E11.5 and E12.5 have a significantly higher number of transcription factor encoding genes up-regulated. Thus, the TEPC transcriptional programme appears to be added onto that of E10.5 3PP cells. Therefore, analysis of E9.5 3PP cells would be useful in determining whether an endoderm transcriptional programme is down-regulated between E9.5 and E10.5. As discussed above, due to the presence of PT cells in the E11.5 samples, some of the genes could be up-regulated in parathyroid cells as opposed to TEPCs. However, such genes would be expected to be significantly down-regulated between the E11.5 and E12.5 as the samples from the later time point should contain far fewer parathyroid cells. Among the 51 transcription factor genes up-regulated at E11.5 compared to E10.5, only 8 were significantly downregulated at E12.5: *Gata3*, *Ebf1*, *Esr1*, *Esrrg*, *Nr4a2*, *Cphx*, *Maf*, and *Mafb*. As discussed above, the expression of *Gata3* appears to be repressed in TEPCs by FOXN1. Thus, enrichment of biological processes related to transcription of the genes up-regulated in E11.5 vs E10.5 comparison seems to represent changes occurring in developing TEPCs. These 51 transcription factors are therefore candidate transcriptional regulators of *Foxn1*. Combining these results with the TFBS analysis in Chapter 4 showed that 11 of the 51 TFs have physical binding sites present in identified *Foxn1* promoter and/or enhancers. Transcription factors such as *Pax1* and *Pax9*, which are likely to be involved in regulation of *Foxn1* expression (unpublished data), were also significantly up-regulated at E11.5 but TFBS analysis did not identify any binding sites within the regulatory elements of *Foxn1*. This could either be due to poor quality of the PWM or these factors could bind to regions outside the putative regulatory elements identified in Chapter 4. Amongst the transcription factors identified, NFATC2 and STAT4 are of particular interest as the genes up-regulated at E11.5 are enriched for genes with binding sites

for these factors in their promoters. NFATC2 is significantly enriched at E11.5, whereas STAT4 is up-regulated by 3.4 fold but is not significant at FDR ≤10% (padj = 0.2). Thus, the role of these two transcription factors should be tested in more detail. NFATC2 also has binding sites within the identified *Foxn1* regulatory regions raising the possibility that NFATC2 might be involved in regulating *Foxn1* in TEC.

The enrichment analysis also identified pathways enriched at the three different developmental stages under study. Interestingly, the E11.5 samples showed up-regulation of MAPK signalling pathway compared to E10.5, suggesting that this pathway might be important for transcriptional changes seen at E11.5. MAPK signalling plays a role in various cellular processes, including proliferation and differentiation, which are both expected of developing thymic rudiment. Thus, this pathway or some of its members are likely to also be involved in proliferation and/or differentiation of the developing thymic rudiment.

The enrichment of NFκB signalling pathway at E12.5 compared to E10.5 suggests that this pathway plays an important role in TEC development. Various signalling pathways influence the activation of NFκB pathway (Figure 5.10) and the results described above suggests its activation is mediated mostly through the binding of IL1β to its receptor. Also, while NFκB is known to be important for mTEC development, its role in TEPC differentiation, if any, remains unexplored. The results from ISMARA suggested that *Nfkb1*, *Rela*, *Irf1*, *Irf2*, and *Irf7* play important roles in E12.5 TEPCs. Furthermore, I showed that the NFκB signalling effector TFs are a part of a transcriptional network involving IRF, STAT, and NFAT TFs, which is likely to be important for TEPC development (Figure 5.14). Finally, I showed that there is a shift from a TF network involving early developmental genes such as *Hox* and *Sox* genes to a network involving *Nfkb* and *Irf* genes (Figure 5.14). Thus, the gene expression changes observed in the RNA-seq data identify genes, processes, and pathways characteristic of each analysed developmental stage and shed new light on the molecular insights of TEPC generation and differentiation.

6. The Role of Smad and TGFβ proteins in regulation of *Foxn1* transcription

6.1 Introduction

The discovery that induced over-expression of *Foxn1* in an aged-involuted thymus can lead to regeneration makes this gene an important target for therapeutic interventions (Bredenkamp, Nowell, et al. 2014). One way to modulate *Foxn1* transcription in TECs would be to target its upstream transcriptional regulators; inhibiting the activity of its transcriptional repressor and/or increasing the activity of its transcriptional activator. However, a lack of understanding of the transcriptional regulation of *Foxn1* in TECs is currently a limiting factor for this approach. An alternate approach to modulating *Foxn1* expression levels is to stimulate or inhibit signalling pathways that influence the level of *Foxn1* transcription. An important advantage of the later approach is that drug molecules capable of inhibiting or activating most of the signalling pathways have already been developed and tested extensively. Therefore, I decided to also focus on identification of signalling pathways that modulate the expression of *Foxn1*.

The TGFβ signalling pathway is important for development, aging, and even cancer (Massagué 2012). The effects of activation of TGFβ pathway are highly cell-type and condition specific. The activation of this pathway has been shown to inhibit proliferation but also to promote cell growth; to enhance pluripotency but also to promote differentiation; to regulate tissue specific transcription factors; and to suppress pre-malignant cells but also to encourage metastasis (Massagué 2012). Furthermore, TGFβ signalling can have either positive or negative effects on transcription of particular genes depending on the target gene and cell type.

There are a large number of TGFβ family proteins, which can broadly be divided into two ligand subfamilies, the TGFβ-activin-Nodal and BMP subfamilies. The TGFβ subfamily consists of three ligands, TGFβ1, TGFβ2, and TGFβ3 and two receptors TGFβRI and TGFβRII. The binding of the ligand assembles a receptor

complex consisting of two TGFβRI and two TGFβRII receptor proteins, following which the type-II receptors (TGFβRII) phosphorylate the type-I (TGFβRI) components. Phosphorylated (activated) type-I receptors subsequently bind to the SMAD proteins, SMAD2 and SMAD3, to phosphorylate and activate them. SMAD2 and SMAD3 are also called R-SMADs (receptor-regulated SMADs) due to their interaction with TGFβ receptors (Massagué 2012). Following activation, two molecules of phosphorylated R-SMAD proteins form trimers with one molecule of SMAD4 (co-SMAD) to form functional units, which are subsequently transported in to the nucleus (Massagué 2012). These activated functional units, and SMAD3 by itself, can bind other transcription factors in the nucleus and are subsequently directed to target genes for transcriptional regulation (Mullen et al. 2011). Various mechanisms exist that target the activated SMAD complexes to promote degradation. Furthermore, activation of TGFβ signalling pathway induces the transcription of *Smad7*, which provides negative feedback regulation to TGFβ signalling. Finally, the TGFβ signalling pathway also interacts with MAPK and PI3K pathways to mediate SMAD-independent effects.

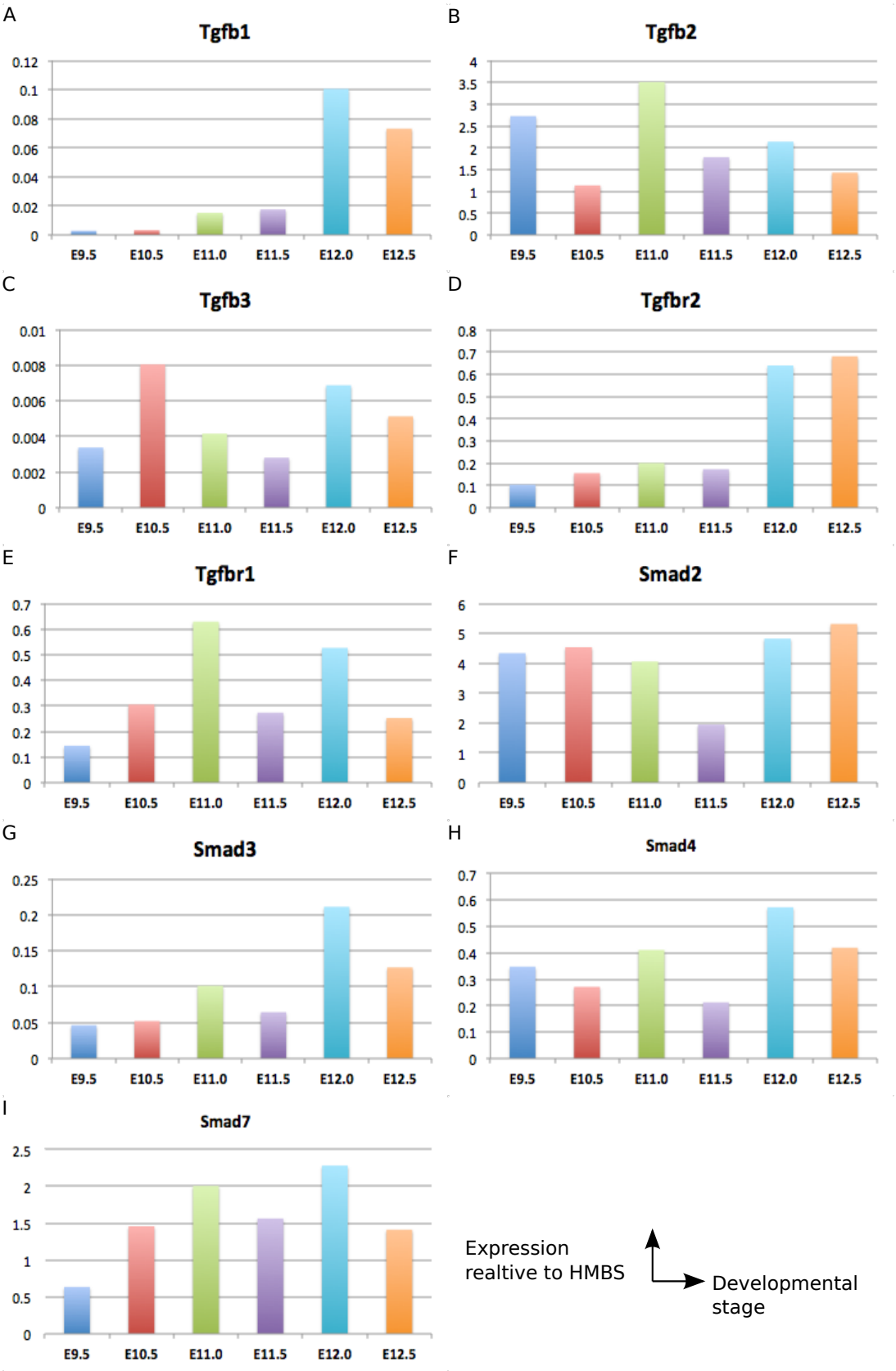
The presence of the pathway terms "TGF-beta signalling pathway", "TGF-beta receptor signalling", "BMP receptor signalling", and the GO Molecular Function terms "transforming factor beta receptor binding", and "SMAD binding" in the analysis of ChIP-seq data shown in Chapter 4 suggested that the TGF-beta and BMP signalling pathways may be important regulators of TECs. Furthermore, Chapter 1 describes the importance of TGFβ and BMP-signalling pathways in the thymus. The slower rate of thymic involution in TGFβ receptor mutant mice suggests a possible interaction between TGFβ signalling and *Foxn1*. Given the important role of *Foxn1* in thymic involution and regeneration, it was desirable to determine whether TGFβ signalling mediates its effect through regulation of *Foxn1* expression in TECs. The identification of SMAD4 binding sites in *Foxn1* promoter, described in Chapter 4, suggested that TGFβ could indeed regulate *Foxn1* expression in TECs. I therefore set out to test this hypothesis and to determine the effects of this pathway on expression of *Foxn1*.

6.2 Expression of TGFβ signalling pathway genes during thymus development

Some of the key genes involved in TGFβ signalling pathway are the three TGFβ ligands (*Tgfb1*, *Tgfb2*, and *Tgfb3*), the two receptors (*TgfbrI* and *TgfbrII*), and the genes encoding for the downstream effector SMAD proteins (*Smad2*, *Smad3*, *Smad4*, and *Smad7*). The RNA-seq data indicated extremely low-level expression of most of these genes in 3PP cells and TEPCs. The expression pattern of these genes was further determined by qPCR. All three TGFβ ligands were expressed in TEPC from as early as E9.5 (Figure 6.1A-C). *Tgfb2* was the highest expressed ligand, with its expression being several hundred folds higher than *Tgfb1* and *Tgfb3*. The expression of *Tgfb1* was significantly up-regulated from E12.0 suggesting that this ligand may play an important role in thymus development after E12.0. Finally, *Tgfb3* was consistently expressed at much lower levels than the other two ligands. *Tgfbr2* showed an expression pattern similar to *Tgfb1*, with its expression being up-regulated at E12.0 (Figure 6.1D). The expression of *Tgfbr1* increased from E9.5 to E11.0 after which its expression changed every 12 hours (Figure 6.1E). However, the expression level of all the TGFβ ligands and receptors, except *Tgfb2*, was low or extremely low (as demonstrated by the scale of Y-axis showing relative expression levels). Interestingly, using RPKM cut-off of 0.3 for the RNA-seq data in Chapter 5 showed an absence of expression for all TGFβ ligands, expect *Tgfb2* at E11.5 and receptors, except *Tgfbr1*. This discrepancy suggests that the RPKM cut-off of 0.3 might be too strict for the RNA-seq data. However, biological replicates for the qPCR based gene

Figure 6.1: Analysis of *Tgfb1*, *Tgfb2*, *Tgfb3*, *Tgfbr2*, *Tgfbr1*, *Smad2*, *Smad3*, *Smad4*, and *Smad7* expression profiles during normal thymus organogenesis.

See following page
Relative expression levels in 3PP and TEP cells were determined by QRT-PCR. Graphs show the expression profiles in WT thymus for (A) *Tgfb1*, (B) *Tgfb2*, (C) *Tgfb3*, (D) *Tgfbr2*, (E) *Tgbr1*, (F) *Smad2*, (G) *Smad3*, (H) *Smad4*, and (I) *Smad7*. Data are shown relative to *HMBS*. Data shown are representative of three technical replicates.


A
Tgfb1
0.12
0.1
0.08
0.06
0.04
0.02
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
B
Tgfb2
4
3.5
3
2.5
2
1.5
1
0.5
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
C
Tgfb3
0.01
0.008
0.006
0.004
0.002
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
D
Tgfbr2
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
E
Tgfbr1
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
F
Smad2
6
5
4
3
2
1
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
G
Smad3
0.25
0.2
0.15
0.1
0.05
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
H
Smad4
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
I
Smad7
2.5
2
1.5
1
0.5
0
E9.5
E10.5
E11.0
E11.5
E12.0
E12.5
Expression realtive to HMBS
Developmental stage

expression data shown in this Chapter (n=1 for data in Figure 6.1) are required to determine the variability of these results, which will allow a more robust comparison with the RNA-seq data.

The downstream SMAD effector proteins are expressed at all time points and their expression showed a spike at E12.0 (Figure 6.1F-H). The negative regulator of TGFβ signalling, *Smad7*, was found to be expressed throughout all the time points analysed and its expression increased from E9.5 to E10.5 after which it showed minor variations in expression (Figure 6.1I). The expression pattern of *Smad7* was similar to that of *Tgfbr1*. The presence of *Smad7* expression suggests that TGFβ signalling is active in TEPCs during development, possibly due to autocrine and/or paracrine signalling by the secreted ligands.

6.3 Increased TGFβ signalling suppresses *Foxn1* expression in *in-vitro* culture of E12.5 thymic cells

The above data, together with inferences from the literature and the identification of SMAD binding sites in the *Foxn1* promoter shown in Chapter 4, suggested that TGFβ signalling might directly affect *Foxn1* expression. To test this, I elected to use short term culture of E12.5 thymic cells. Thus, E12.5 thymi were isolated by microdissection. The dissected thymi were dissociated to form a single cell suspension and were plated on to a thin layer of ECM in N2B27 medium (Figure 6.2A). The cells were then cultured for 72 hours in presence or absence of TGFβ1 or A8301, a TGFβRI kinase inhibitor, following which $EpCAM^+$ TEPCs were sorted for RT-qPCR analysis.

The cells cultured in the presence of 10μg TGFβ1 showed a significant reduction in *Foxn1* expression (p-value = 0.008), which was rescued in the presence of 500nM of the TGFβ inhibitor A8301 (p-value for difference from N2B27 = 0.19) (Figure 6.2B). This suggested that TGFβ signalling negatively regulates the expression of *Foxn1* in TEPCs, at least in *in-vitro* cultures. The expression of *Smad7* was up-

regulated in the presence of TGFβ1 but not in its absence or in the presence of A8301 (Figure 6.2C). The reduction in *Foxn1* expression was also associated with reduction in functional FOXN1 protein, as inferred by the reduction in the expression of *Dll4* (p-value = 0.002) (Figure 6.2D). *Pax1* also showed similarly decreased expression in the presence of TGFβ (p-value = 0.02) (Figure 6.2E). This is consistent with the high correlation between expression of *Pax1* and *Foxn1* described in previous chapters and further supports the hypothesis that these two genes are a part of the same transcription factor network in TEPCs.

On the other hand, the expression of *Tbx1* was increased in the presence of TGFβ (p-value = 0.03) (Figure 6.2F). As mentioned previously, ectopic *Tbx1* expression in developing thymic rudiment results in loss of *Foxn1* expression (Reeh et al. 2014). Thus, the observed decrease in *Foxn1* expression could result from a direct effect of TGFβ signalling, or an indirect effect of the signalling pathway on *Pax1* and/or *Tbx1*. Furthermore, the presence of TGFβ1 in culture also increased the expression of *Gata3* (Figure 6.2G). Together, the changes in expression of the above mentioned genes suggests that TEPCs shift to a developmentally earlier transcriptional program in the presence of TGFβ1 in culture. An alternate possibility is that TGFβ1 could have a selective effect on the cultures, supporting parathyroid cells while leading to a loss of TECs. Analysis of expression of *Gcm2* and/or *Pth* or that of parathyroid-specific cell-surface marker through flow cytometry could help address this issue.

6.4 Inhibition of TGFβ signalling in-vivo results in increased *Foxn1* expression in several TEC subsets

Hauri-Hohl et al. recently showed that inhibition of TGFβ signalling *in-vivo* resulted in a numerical and proportional increase in mTEC (Hauri-Hohl et al. 2014). However, the interaction of this signalling pathway with *Foxn1* was not investigated. To address this issue, I injected 4 weeks old C57BL/6 (BL6) mice with either A8301 (0.5mg/kg body weight) or vehicle only for 7 consecutive days (Figure 6.3A). The

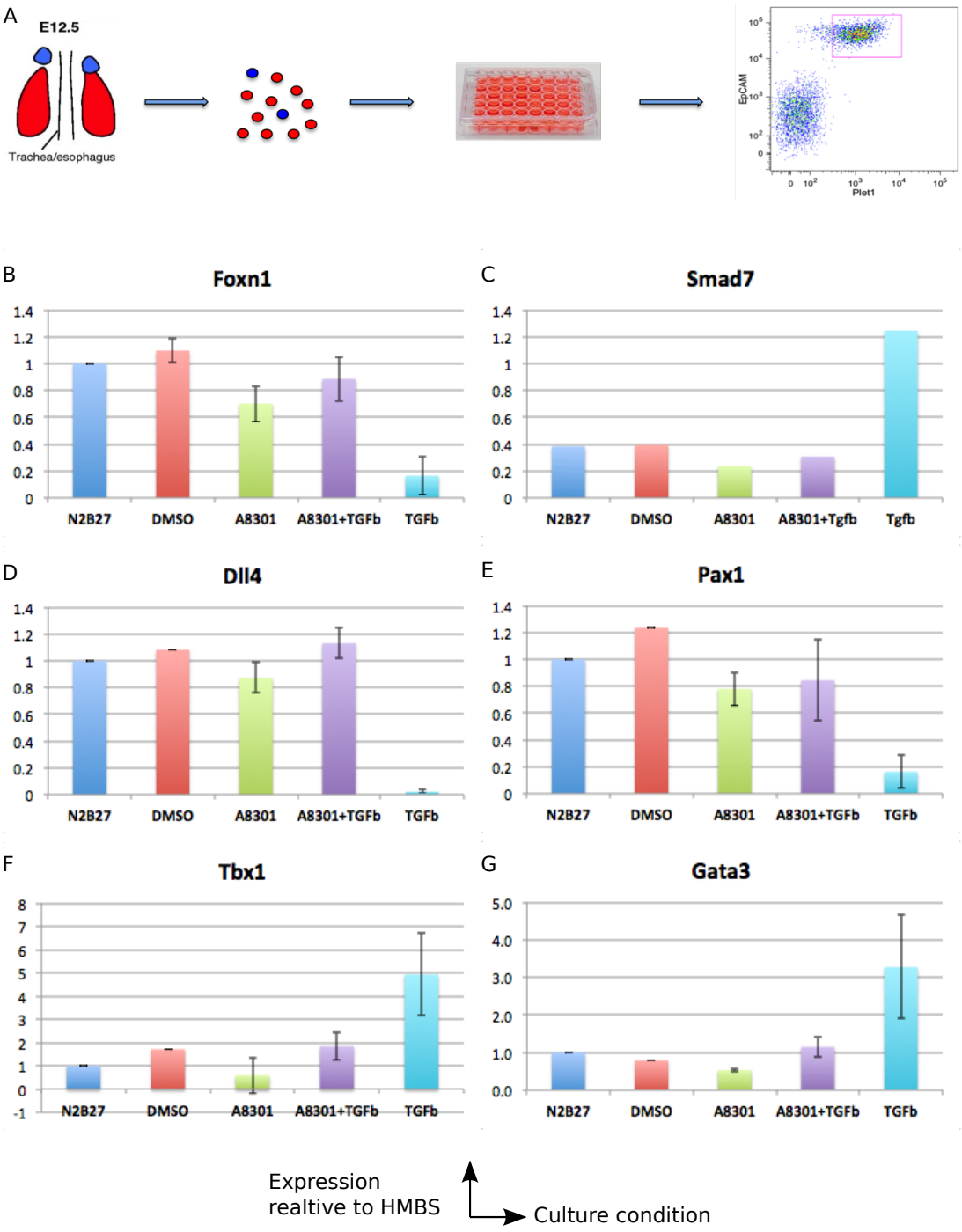


Figure 6.2: Effects of TGFβ signalling on thymic epithelial cells culture.
(A) Schematic of the experimental setup used for testing the effects of TGFβ signalling on thymic epithelial cells culture. E12.5 thymi were isolated and dissociated into single cells suspension. Cells were then cultured on a thin layer of matrigel in N2B27 media in the presence or absence of TGFβ ligand and/or inhibitor. $EpCAM^{+}$ cells from the culture were sorted after 72hours for gene expression analysis. (B), (C), (D), (E), (F), and (G) Analysis of gene expression for (B) *Foxn1*, (C) *Smad7*, (D) *Dll4*, (E) *Pax1*, (F) *Tbx1*, and (G) *Gata3* in thymic epithelial cells cultured in the presence or absence TGFβ ligand and/or inhibitor. DMSO was used as carrier control. Data are shown relative to HMBS. Data shown are representative of at least two biological replicates and three technical replicates. Error bars represent standard deviation.

injected mice were then sacrificed at either day8 or day14 after the first injection, to analyse the effect of inhibition of TGFβ signalling.

The mice injected with vehicle only gained weight over the duration of injections. Their counterparts injected with A8301 also gained weight, but at much lower rate (Figure 6.3B). Thus, inhibition of TGFβ signalling had a systemic effect on mice. On the other hand, the mouse injected with A8301 had a slightly higher thymus weight and higher thymus vs body weight ratio at day 8 compared to that injected with vehicle only (Figure 6.3C&D). Both mice sacrificed at day-14 had gained similar amounts of weight since day-8, regardless of whether they were injected with A8301 or vehicle only, suggesting that the observed effect of inhibiting TGFβ on body weight is reversible. The mice sacrificed at day 14 also did not show any substantial difference in thymus vs body weight ratio (Figure 6.3C). A pharmacological inhibition of TGFβ signalling leads to increased mTEC cellularity (Hauri-Hohl et al. 2014), however, its effect on thymocytes has not been demonstrated. A T-cell specific deletion of *Tgfbr1* leads to defects in specification of $CD8^+$ T-cell fate and reduced frequency of $FoxP3^+$ Treg (Liu et al. 2008; Ouyang et al. 2010; Ouyang et al. 2013). This suggests that pharmacological inhibition of TGFβ signalling could affect the thymocyte numbers and development. Thus, the effect of TGFβ inhibition on thymus weight observed here represents a cumulative effect on TECs and thymocytes.

Figure 6.4A shows the sorting strategy for isolation of TEC subsets. The proportions of these TEC sub-populations were similar in both A8301 treated and vehicle treated mice analysed at day 8 and day 14 after the first injection (Figure 6.4B) (n = 1). While small variations were observed in proportions of cTEC $MHCII^{hi}$ and $Ly51^+Plet1^+$ cells between A8301 treated and control mice, more biological replicates are required to determine whether these differences are significant.

At gene expression level, the expression of *Foxn1* was higher in mTEC $MHCII^{hi}$, cTEC $MHCII^{hi}$, and cTEC $MHCII^{low}$ at day 8 in the mouse injected with A8301

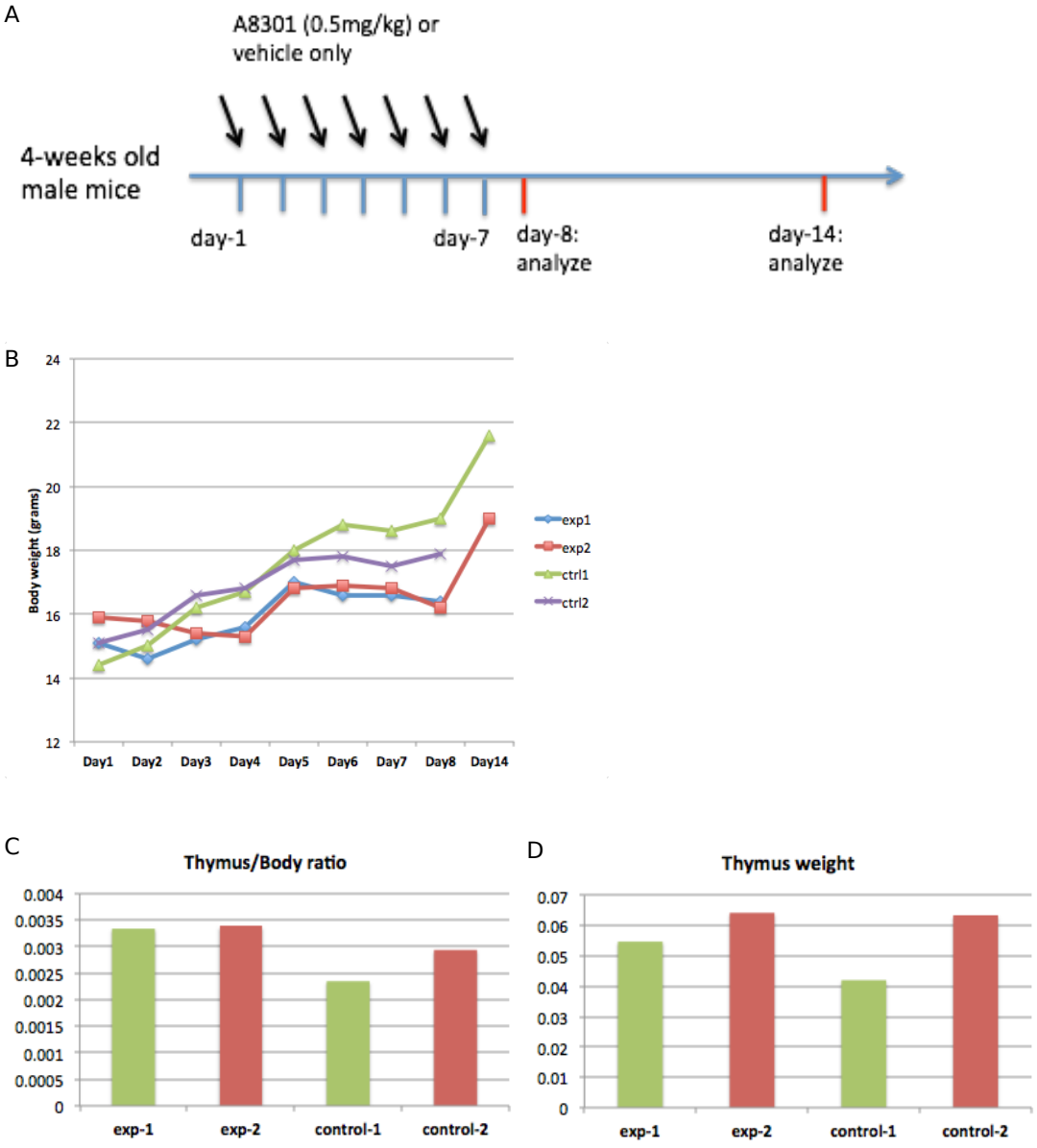


Figure 6.3: Analysis of the effects of TGFβ signalling inhibition on adult thymus
(A) Schematic of the experimental setup used for testing the effects of inhibiting TGFβ signalling in adult thymus. (B) Comparison of changes in body weights of A8301 and vehicle only treated mice. (C) and (D) Bar graphs comparing the (C) thymus/body weight ratio and (D) thymus weight of A8301 and vehicle only treated mice at the end of the experiment period. The mice represented as control-2 and exp-1 were analysed 8 days after the first injection, while the mice represented as control-1 and exp-2 were analysed 14 days after the first injection.

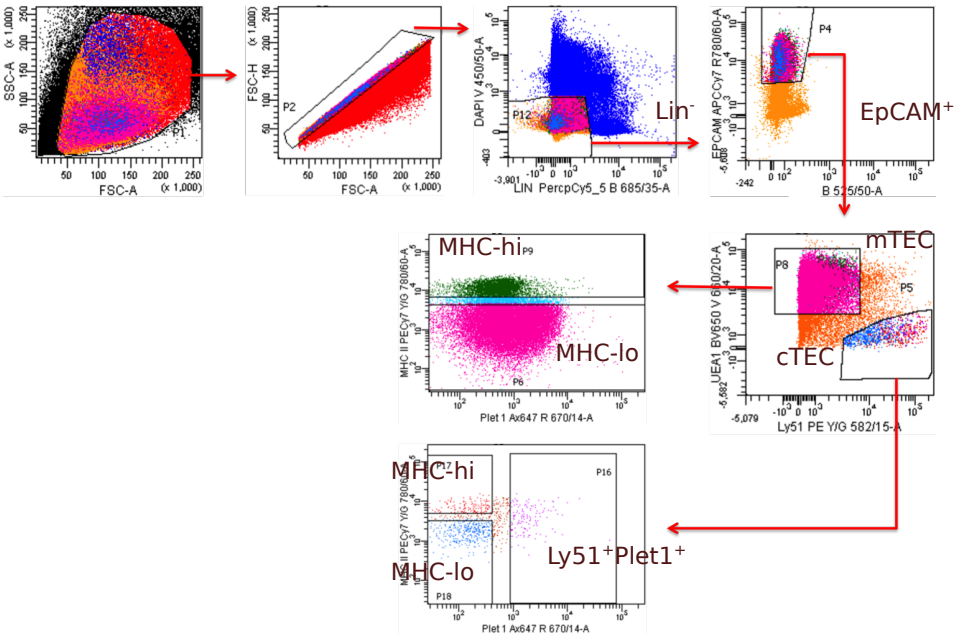
compared to the vehicle only control (Figure 6.5A). Interestingly, *Foxn1* expression in cTEC MHCII^{hi} was lower in A8301 treated mice compared to control at day 14, suggesting that the upregulation of *Foxn1* expression observed at day 8 is reversible (Figure 6.5A`). The Ly51^{+}Plet1^{+} population, which contains thymic epithelial stem cells (Ulyanchenko et al. unpublished) showed a decrease in the expression of *Foxn1* at both day 8 and day 14 in the mouse injected with A8301 compared to the control. *Foxn1* was expressed at similar levels between A8301 treated and control mice at day-14 (Figure 6.5A-A`). Thus, the expression of *Foxn1* differed between A8301 treated and control mice; however, more biological replicates are required to test the reproducibility and significance of these results.

An increase in the expression of *Foxn1* in the adult thymus has been shown to result in increased expression of genes such as *Dll4*, *Ccl25*, and *Pax1* (Bredenkamp, Nowell, et al. 2014), similar to that observed during thymus development. Interestingly, the expression of these genes was either unchanged (or down-regulated in the case of *Pax1*) in A8301 treated mice compared to controls. (Figure 6.5B-B` to Figure6.5D-D`). This is consistent with the previous observation by Jeker and colleagues that loss of *Smad4* in TECs results in an increase in *Foxn1* expression in certain TEC subpopulations but a decrease in the expression of *Ccl25* (Jeker et al. 2008). Whether the observed increase in *Foxn1* expression results in an increase in FOXN1 protein levels remains to be determined and could help understand the discrepancies in the observed results. Furthermore, another possible factor influencing the results is the age of the mice. Investigation of the outcome of inhibiting TGFβ signalling in older mice, which have a more involuted phenotype, would indicate whether age (and therefore the state of involution) determines the response to TGFβ inhibition.

6.5 Discussion

In this chapter, I show that inhibition of TGFβ signalling in the thymus leads to an increase in the expression of *Foxn1*. I further show that the negative effect of TGFβ

A



B

Percentage of EpCAM+ cells					
	mTEC MHC-hi	mTEC MHC-lo	cTEC MHC-hi	cTEC MHC-lo	Ly51+Plet1+
control (day-8)	11.65	71.04	0.42	1.75	0.53
A8301 (day-8)	13.62	67.83	0.5	1.4	0.34
control (day-14)	10.51	42.98	0.89	0.42	0.28
A8301 (day-14)	9.39	47.04	1.16	0.39	0.45

Figure 6.4: Sorting strategy for isolation of cTECs and mTECs from adult thymus.
(A) Gates showing isolation of MHC Class II^{hi} and MHC Class II^{lo} cTECs and mTECs using flow cytometry. Events recorded as single cells were gated for cells negative for dead cell marker DAPI and Lin (CD45, CD11b, CD31, Ter119, Pdgfa). EpCAM^{+} cells within this population were gated on UEA1 and Ly51 to identify mTECs and cTECs respectively. mTECs were further gated on MHC Class II expression to obtain mTEC MHC-hi and mTEC MHC-lo populations. cTECs were divided based on presence or absence of Plet1 and Plet1^{-} population was further gated on MHC Class II to obtain cTEC MHC-hi and cTEC MHC-lo populations. (B) Table showing the percentage of each mTEC and cTEC subpopulations within the EpCAM^{+} population in A8301 and vehicle treated mice.

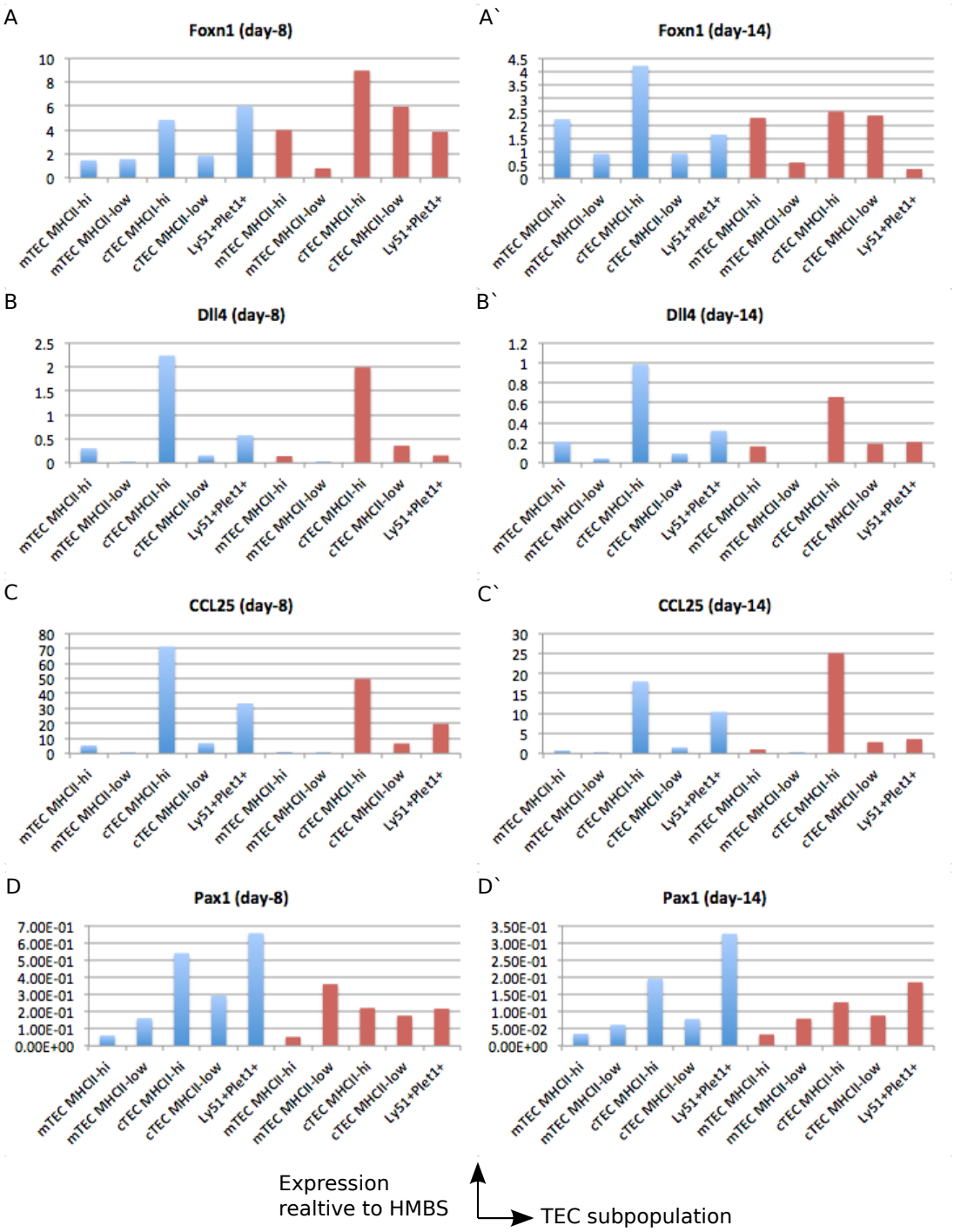


Figure 6.5: Analysis of gene expression for *Foxn1*, *Dll4*, *Ccl25*, and *Pax1* in TEC subpopulations in mice treated with A8301 or vechicle only.

Analysis of gene expression for (A) *Foxn1*, (B) *Dll4*, (C) *Ccl25*, and (D) *Pax1* in mTEC MHC-hi, mTEC MHC-lo, cTEC MHC-hi, cTEC MHC-lo, and Ly51^{+}Plet1^{+} cells isolated from mice treated with either A8301 or vehicle only. Graphs on the left and right represent results from mice sacrifised 8 and 14 days after the first injection, respectively. The blue bars represent vehicle only treated mice whereas the maroon bars represent A8301 treated mice. Data are shown relative to HMBS. Data shown are representative of three technical replicates.

signalling on *Foxn1* expression is at least partially mediated through increased expression of *Tbx1* in TECs.

6.5.1 TGFβ signalling during thymus development

The TGFβ family of proteins play important roles in the development and maintenance of many cell types and tissues. However, their importance in regulation of the fetal and postnatal thymus has not been determined in detail. Among the various signalling pathways orchestrated by TGFβ proteins, the roles of BMP and TGFβ signalling pathways have been studied to some extent in thymic epithelial cells. In this chapter, I investigated the effects of TGFβ signalling on *Foxn1* expression in the developing and adult thymus. These data show that inhibition of TGFβ signalling pathway increases *Foxn1* expression in TEPCs and some postnatal TEC subsets.

Previous studies analysing the role of TGFβ in thymus have used $Foxn1^{Cre}$ to delete TGFβ receptors in the thymus, which results in deletion of the receptor upon activation of the *Foxn1* gene at around E11.5 (Hauri-Hohl et al. 2008; Odaka et al. 2013; Hauri-Hohl et al. 2014). These transgenic mice models therefore did not address the role of TGFβ signalling in thymus development prior to *Foxn1* expression, including in regulation of the initiation of high-level *Foxn1* transcription. The results from above analyses suggest that down-regulation of TGFβ signalling activity might be required for normal thymus development. This hypothesis could be tested by using $Foxg1^{Cre}$, which is expressed in 3PP from E9.5, to delete *Tgfbr1/2*.

It has been shown that a short pulse of BMP signalling is required for initiation of *Foxn1* expression during thymus development (Neves et al. 2012), however whether a similar temporal modulation of TGFβ signalling is also important remains to be determined and can not be answered using the above mentioned mouse models. Null mutants of either *Tgfbr1* or *Tgfbr2* exhibit defects in yolk sac, placenta, and various defects in hematopoetic and immune systems and are thus not suitable for studying

the role of TGFβ signalling in thymus development. Thus, new mouse models targeting deletion of either *Tgfbr1* or *Tgfbr2* specifically in the 3PP endoderm and TEPCs prior to initiation of *Foxn1* expression are required to address the above issues.

After the initiation of *Foxn1* expression, the deletion of *Tgfbr2* does not appear to have any adverse effect on thymus development or function (Hauri-Hohl et al. 2008; Hauri-Hohl et al. 2014). This suggests that TGFβ signalling is not required for *Foxn1*-dependant thymus development. However, I here show that addition of TGFβ1 to *in-vitro* cultures of developing E12.5 thymic lobes suppresses *Foxn1* expression relative to controls. It is important to note that the culture system used here dissociates the 3D E12.5 thymic rudiment and thus represents an artificial system, which may not resemble the *in-vivo* organ. Thus, whether TGFβ signalling has a similar effect on *Foxn1* expression during thymus development *in-vivo* remains to be determined. This question could be approached by administration of a TGFβ signalling inhibitor to pregnant mice, however the outcome would depend on good availability of the inhibitor within the developing embryo. One approach to ensure good availability of the inhibitor to the developing embryo would be in-utero injections. Another possible *in-vitro* approach could employ whole embryo cultures in presence or absence of TGFβ1 to address the above issues. The above results suggest that induced activation of TGFβ signalling in the developing 3PP would result in ectopic expression of *Tbx1* and prevent the up-regulation of *Foxn1* expression in the thymus domain.

6.5.2 The importance of TGFβ signalling during thymus homeostasis and involution

Thymic involution is an important aspect of thymus biology and it is desirable to improve our understanding of this process to enable development of therapies to improve thymus function in some cases of immunocompromised patients. The TGFβ signalling pathway has been shown to influence the rate of thymic involution (Hauri-Hohl et al. 2008) and the distribution of TEC subsets and thymus function in adult

mice (Hauri-Hohl et al. 2014). The preliminary results presented herein, in which administration of a TGFβ signalling inhibitor was used to inhibit TGFβ signalling, supported the hypothesis that TGFβ signalling could regulate *Foxn1* expression in TECs. The inhibition of TGFβ signalling using a pharmacological inhibition had a positive effect on *Foxn1* expression in some TEC subsets. However, more replicates are required to determine the reproducibility and significance of the results shown above. Furthermore, as the mice used in these experiments were 4-weeks old, and had therefore not commenced thymic involution, they might be less responsive to the inhibition of TGFβ signalling pathway. Thus, similar A8301 administration experiments should be carried out in older mice. Finally, a thorough analysis should be carried out using different doses of A8301 inhibitor, in order to determine the optimal dose that shows a definitive phenotype without any adverse side effects.

Together, the data presented in this chapter indicate the potential importance of TGFβ signalling in regulation of *Foxn1* expression in TEPCs and TECs. Whether this regulation is mediated through SMAD proteins or through other associated pathways, such as p38 and MAPK pathways, remains to be determined. Regardless, modulation of TGFβ signalling could represent a novel mechanism for pharmacologically regulating *Foxn1* expression in TEPCs and TECs. Whether such pharmacological intervention can produce sufficient up-regulation of *Foxn1* expression in TECs to lead to partial or complete thymus regeneration remains to be determined. It is likely that modulation of TGFβ pathway in combination with other signalling pathways, such as BMP, would be required to produce clinically relevant thymus regeneration through up-regulating the expression of *Foxn1*.

7. Discussion

The field of stem cells has seen massive progress over the last few years in bringing regenerative and cellular therapies closer to clinics. A crucial factor in this success has been the improvement in our understanding of the development of different cell types, tissues, and organs. The thymus plays an important role in the adaptive immune system and it is therefore desirable to generate therapies for boosting its function in certain sub-groups of immunocompromised patients and potentially, the elderly population. Recent advances, including in differentiation of pluripotent stem cells and in reprogramming of MEFs to TEC-like cells, point to the capacity to provide cells grown *in vitro* for cellular therapies, aimed at boosting the thymus *in vivo* and are therefore major advances in this field (Parent et al. 2013; Bredenkamp, Ulyanchenko, et al. 2014). Furthermore, the study showing complete thymus regeneration in aged mice is the first of its kind showing the complete *in vivo* regeneration of an organ using single transcription factor (Bredenkamp, Nowell, et al. 2014). Thus, several different approaches are being investigated for boosting thymus function in patients. Importantly, the transcription factor FOXN1 is essential and sufficient for both the *in vivo* regeneration and *in vitro* reprogramming mentioned above, and is thus considered a master regulator in TECs. In this thesis, I set out to investigate the transcriptional regulators of *Foxn1* in TEPCs, with the aim of improving the current understanding of the regulation of *Foxn1* expression in TEPCs and TECs. The results from this thesis are summarised graphically in Figure 7.1.

The work described in Chapter 3 explored the transcription profiles for the candidate transcriptional regulators that had been predicted in a bioinformatics analysis performed prior to the start of this PhD. This analysis had sought to identify genes whose expression pattern correlated with that of *Foxn1* in E12.5 and E15.5 TEPCs and E15.5 TECs, in an attempt to identify potential *Foxn1* regulators. Thus, it was important to study the expression patterns of these genes during earlier thymus development, prior to and during the onset of *Foxn1* expression. A transcriptional activator of *Foxn1* would be expected to be expressed before, or at least immediately

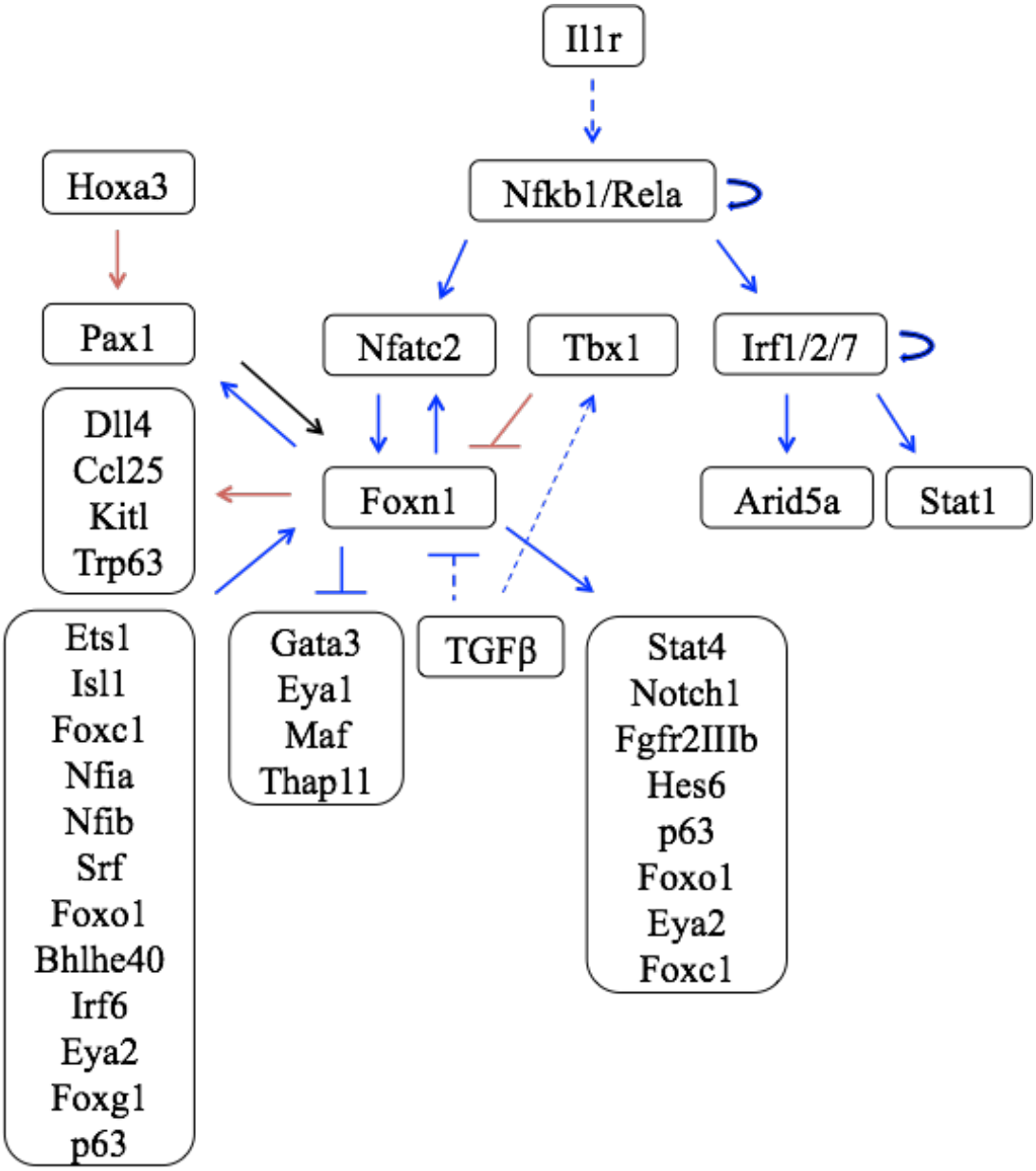


Figure 7.1: Predicted genetic interactions in TEPCs.
Shown are the genetic interactions predicted in Chapters 3 to 6 are shown as a network. Also shown is potential regulation of *Foxn1* and *Tbx1* expression by TGFβ signalling. The edges of the network are colour coded to represent the source of the predicted interactions: blue - this thesis, black - unpublished data from Blackburn lab, and brown - literature. Inputs from signalling pathways into this network are shown usign dashed lines.

prior, to the onset of *Foxn1* expression in TEPCs. Chapter 3, therefore, analysed the expression of the candidate regulators during 3PP and TEPC development in both WT and *nude* (i.e. *Foxn1* null) mice. Most of the genes studied were expressed before or at the time of onset of *Foxn1* expression. Comparison between WT and *nude* further showed that the expression of some of these genes was sensitive to the presence or absence of FOXN1. These analyses allowed identification of genes whose expression was dependent on FOXN1, such as *Gata3*, *Eya1* and *Thap11*, which were repressed by FOXN1. *Foxn1* has been shown to be important for the expression of *Notch1* in hair follicle (Cai et al. 2009). *Notch1* expression was found to be upregulated substantially at E12.5, a time point at which highest expression of *Foxn1* was observed, suggesting that FOXN1 regulates the expression of *Notch1* in TECs, similar to that observed in hair follicle. The expression of *Hes6*, an effector of Notch signalling, appeared to be regulated by *Foxn1*, suggesting that *Foxn1* might be involved in modulating Notch signalling pathway in TECs. Furthermore, these analyses also identified genes whose expression pattern during early thymus development was consistent with that expected for a transcriptional regulator of *Foxn1*, such as *Pax1*, *Foxo1*, *p63*, *Eya2*, *Hes1*, *Bhlhe40*, and *Irf6*. Indeed, unpublished data from this lab suggest that *Pax1* and *Pax9* are important for initiation of *Foxn1* expression in the fetal thymus. However, the expression of some of these genes, such as *Pax1*, *Foxo1*, *Eya2*, and *p63*, was reduced in *Foxn1*$^{-/-}$ TEPCs, suggesting that these genes could also be targets of FOXN1. It is possible that some of these genes could act as both transcriptional regulators and targets of *Foxn1*, similar to the observations for *Pax1*. Another set of genes, including *Pax9*, *Foxg1*, *Sox9*, and *Six1*, showed an increase in expression at E12.0. Whether such an increase in the expression of these genes is required for upregulation of *Foxn1* expression at this developmental stage remains to be determined.

The results presented in this chapter suggested that the expression of *Pax1* in TEPCs is positively regulated by FOXN1, whereas *Pax9* expression was found to be similar between WT and *nude* thymi. Thus, *Pax1* and *Foxn1* can act as positive regulators of each other. *Foxn1* has also been suggested to be show positive auto-regulation (Zook et al. 2011; Bredenkamp, Nowell, et al. 2014). Thus, there seems to be a highly

stable transcription factor network prevalent in TECs with several positive feed-forward mechanisms strengthening the TEC transcriptional programme. The downregulation of *Eya1* and an upregulation of *Pax1* from E11.5 to E12.5 suggest that *Foxn1* could replace *Eya1* in regulation of *Pax1*. Since *Eya1* is required for 3PP development, an interesting question is whether a 3PP developmental transcription programme is being replaced by a TEC development and differentiation programme. Together, the results from this chapter identified potential *Foxn1* target genes and genes with expression profiles consistent with that expected for transcriptional activators of *Foxn1*. However, an important limitation of this approach was the correlation analysis used for identification of candidate transcriptional regulators of *Foxn1*. As mentioned above, the correlation analysis was performed on a small number of samples (5 samples) from developmental time points post initiation of *Foxn1* expression in TEPCs and thus had limited statistical power. Furthermore, unlike *Pax1* and *Pax9*, our knowledge regarding the functions of most of the candidate transcriptional regulators in the thymus is very limited. It was therefore desirable to complement this approach with genome-wide unbiased approaches for identification of transcriptional regulators of *Foxn1*.

To this end, Chapter 4 focused on identification of regulatory elements governing the expression of *Foxn1* in developing TEPCs. The approach selected for this purpose was identification of promoter and enhancer related histone modification through ChIP-seq. An advantage of ChIP-seq is that it allows identification of regulatory elements without a need for any prior knowledge of regulatory interactions between genes. The aim here was to identify small, defined regions that could act as promoter or enhancers to regulate *Foxn1* expression in TEPCs. The commonly used 27 kb *Foxn1* promoter region, which recapitulates *Foxn1* expression in the fetal thymus, contains thousands of binding sites for hundreds of transcription factors and is thus unsuitable for identification of candidate transcriptional regulators. Thus, I reasoned that smaller regions identified using the above ChIP-seq approach would be better suited for this purpose. To this end, I identified five putative enhancers for *Foxn1*, each only a few hundred base pairs long, and the TFBSs present within these regions. Among the genes analysed in Chapter 3, *Ing4*, *Foxa2*, *Hes1*, *Foxc1*, *E2Fs*, and *p53*

had physical binding sites within the identified *Foxn1* promoter or enhancers, further supporting their involvement in regulating *Foxn1* expression. However, whether these genes bind *in vivo* to the identified sites remains to be determined. Analysis of the identified promoter regions for *Foxn1*, *Pax1*, *Pax9*, *Trp63*, *Dll4*, *Ccl25*, *Kitl*, and *Gata3* showed that binding sites for BRCA1:USF2 complex were enriched in these regions. The role of *Brca1* in TECs remains unknown, however the present analyses suggests that it could be involved in regulating gene important for thymus development and function.

A limitation of this approach is that position weight matrices for most of the transcription factors studied here are derived from *in-vitro* or sequence alignment approaches, which are typically of poorer quality than those derived from ChIP-seq data for transcription factors. This is likely why no binding sites for FOXN1 were identified within the regulatory regions of genes thought to be its direct targets, such as *Dll4* and *Ccl25*. Furthermore, it will be important to determine the reproducibility of the identified peaks for histone modifications using biological replicates.

The analysis of the identified genome-wide active enhancer regions revealed that these regions are present in the vicinity of genes important for thymus development and function, supporting this approach. Enrichment analysis performed on these genes showed that enrichment for several signalling pathways, including TGFβ, BMP, Notch, NFκB and MAPK signalling. The TFBS analysis showed presence of several SMAD4 binding sites in the *Foxn1* promoter and enhancers identified herein. Given the importance of both BMP and TGFβ signalling pathways in TECs, it is likely that SMAD proteins may be involved in regulating *Foxn1* expression in these cells. The enrichment of Notch signalling from this analysis was consistent with the importance of this pathway in thymus development and its modulation by *Foxn1* predicted in Chapter 3. The enrichment analysis also suggested that some of the genes regulated by *Foxn1* maybe be shared between the thymus and the skin.

In Chapter 5, I performed global transcriptome analysis on E10.5 and E11.5 3PP cells and E12.5 TEPCs, with the aim of identifying differences in gene expression between these cells. This chapter overcomes the limitations associated with the correlation analysis mentioned above. Furthermore, analysis of the developing 3PP and TEPCs at the selected time points could identify genes and signalling pathways important for this process. The results in this chapter showed that genes which are expressed at higher levels in E12.5 TEPCs compared to the 3PP cells were enriched for biological processes associated with T-cell development and differentiation, processes mediated by cross-talk between thymocytes and TECs. A closer analyses of these genes suggested that they were likely to be important for the antigen presentation function of TECs. On the other hand, genes with decreasing expression from E10.5 to E12.5 were enriched for biological processes related to development, cell fate commitment, and others, supporting the developmentally earlier state of these cells compared to E12.5 TEPCs. The above analysis identified genes that could be important for 3PP endoderm, such as *Nkx2.6* and *Sox2* and also those important for TEC function such as *Cd83*, *Itgb2* and *Prss16*. The roles of these genes in the respective cell types remains to be determined.

I also identified transcription factors expressed differentially between E10.5 and E11.5 3PP cells and showed that of these, *Ets1*, *Gata3*, *Isl1*, *Ebf1*, *Egr1*, *Foxc1*, *Nfia*, *Nfib*, *Nfatc2*, *Srf*, and *Maf* had physical binding sites in the *Foxn1* promoter or enhancers identified in Chapter 4. Thus, these TFs could be involved in regulation of *Foxn1* expression in TEPCs. Some of these TFs were among the list of candidate transcriptional regulators studied in Chapter 3. Of the TFs with significantly upregulated expression at E11.5, *Nfatc2* and *Stat4* are strong candidate transcriptional targets of FOXN1. Furthermore, these TFs may also be involved in regulating *Foxn1* expression as the *Foxn1* promoter and enhancers identified in Chapter 4 contained binding sites for these TFs. The results presented in this chapter suggest that *Nfatc2* could be involved in regulating a large array of genes in TEPCs and thus play an important role in their development and/or differentiation. Finally, analysis of signalling pathways suggested that NFκB signalling is likely to be important for TEPC development and differentiation, consistent with the enrichment

of this pathway in analysis performed in Chapter 4. A closer inspection of this pathway suggested IL1R1 as the major activator of this pathway in TEPCs. Furthermore, the NFκB signalling pathways appears to be regulating genes involved in cell survival and lymphocyte attraction and activation in TEPCs. Genes involved in negative feedback of NFκB signalling were also upregulated suggesting that the activation of this pathway may similar oscillatin patterns as observed in other cell types.

The analysis of predicted interactions between TFs showed that the activity of a TF network involving *Hox* genes decreased from E10.5 to E12.5, probably as a result of negative regulation via *Hoxa7*, whereas the activity of a TF network involving *Nfkb* and *Irf* genes increased during these developmental stages. This further supports the importance of NFκB signalling in these cells. Finally, *Msx2* has been suggested to be upstream of *Foxn1* expression in hair follicle, as *Foxn1* expression is down-regulated in *Msx2* mutants (Meier et al. 1999; Schlake & Boehm 2001). However, *Msx2* expression decreased from E10.5 to E12.5 in the developing thymic primordium, suggesting that *Msx1* does not have a similar role in TEPCs.

In Chapter 6, I investigated the role of TGFβ signalling in regulation of *Foxn1* transcription in TEPCs and TECs. The results from Chapter 4 strongly suggested a role of this signalling pathway in TEPC, while evidence from literature suggest that this pathway is important for thymus homeostasis and involution. Given the important role of *Foxn1* in thymus homeostasis and involution, I wanted to determine whether the TGFβ signalling could influence the levels of *Foxn1* transcription in TECs. The results presented in this chapter suggested that TGFβ signalling could influence levels of *Foxn1* expression both *in-vivo* and *in-vitro*, with inhibition of the signalling pathway resulting in increased expression of *Foxn1*. TGFβ signalling was also found to be able to influence the expression of *Pax1*, *Tbx1*, and *Gata3*. Of these, *Pax1* and *Tbx1* are known to be able to regulate *Foxn1* expression in TEPCs and *Gata3* was found to be a *Foxn1* target in Chapter 3. This suggested that the TGFβ signalling could influence the genes involved in this genetic

network. However, whether the effect of TGFβ signalling on *Foxn1* expression is a direct or indirect effect remains to be determined. Given that *Foxn1* promoters and enhancers identified in Chapter 4 contained SMAD4 binding sites, it suggested that the observed effects on *Foxn1* expression could be a direct effect of the pathway. This hypothesis needs further testing through manipulation of the identified SMAD4 binding sites. To this end, I generated ES-cell lines in which the regions containing SMAD4 binding sites have been deleted. Chimeric embryos generated using these cell lines could help determine the importance of SMAD4 in regulating *Foxn1* expression. Given that SMAD4 is also important for BMP signalling, such chimeric embryos could also be used to determine whether the role of BMP signalling in initiation of *Foxn1* expression is dependant on SMAD4.

7.1 Future experiments

The data presented in Chapter 4 were obtained through analysis of a single sample for each histone modification. It will therefore be important to determine the reproducibility of these results after incorporation of biological replicates into the study; this would allow identification of high-confidence reproducible peaks for histone modifications under study. The identified regulatory regions should also be tested for their respective promoter or enhancer activities using a combination of *in vitro* (for example, Luciferase reporter assays) and *in vivo* approaches (for example, deletion of these regions) to determine the effect on target gene expression during thymus development. The binding sites for the candidate regulators of *Foxn1* shown in Figure 7.1 should be tested using the above approaches in order to determine which of these TFs are involved in regulation of *Foxn1* expression in the thymus. To this end, I have generated an ES cell lines in which the SMAD4 binding sites in the *Foxn1* promoter have been deleted (see Appendix for details). These cell lines should be used to generate high contribution chimeras, an analysis of which would indicate the effect of deleting these TFBSs on *Foxn1* expression during thymus development.

Furthermore, studying the epigenetic changes associated with TEPC development could help identify important regulatory regions that govern the transcriptional

changes associated with this process. It would, therefore, be important to perform ChIP-seq for histone modifications on E10.5 and/or E11.5 3PP cells to carry out a comprehensive analysis of the epigenome of developing TEPCs. This indeed presents additional challenges, as the E10.5 and E11.5 primordium are considerably smaller than the E12.5 thymic rudiment and thus the cell numbers would be even more limiting. An alternative approach, discussed in Chapter 4, would be to characterize the open-chromatin profiles of 3PP cells and TEPCs using ATAC-seq. The combination of the histone modification data generated here and corresponding ATAC-seq data for E12.5 TEPCs could help identify regulatory regions with high confidence.

In Chapter 5, it would be important to test the predicted interactions between TFs using *in vivo* and *in vitro* approaches. The expression of *Nfatc2* and *Stat4* in the developing thymic rudiment should be validated using qRT-PCR, *in-situ hybridization* or immunohistochemistry. It would be important to test the predicted transcriptional regulators of *Foxn1* through deletion of their binding sites, as described above, or through use of conditional mutants for these genes. The importance of NFκB signalling in TEPC development could be tested through deletion of conditional *Nfkb1* and *Rela* alleles using *Foxg1*-Cre, which in 3PP is only active in the endodermal cells. To further investigate the predicted interactions between TFs, it would be interesting to determine whether these TFs exhibit overlapping expression in the thymus domain during organogenesis. Direct targets of these TFs could be identified using ChIP as ChIP-grade antibodies for some of these are commercially available.

The transcription profile for TEPC generated here could be used to determine the relation between TEPCs and more differentiated TEC subpopulations found in the adult thymus through comparison of global trancscriptome data. Transcriptome data for adult TEC subpopulations was recently generated in our lab (Kathy O'Neil, unpublished) and could be used for this purpose. The regulatory networks governing the early development of blood tissue has been identified using gene expression

profiling of hundreds of single cells (Moignard et al. 2013; Moignard et al. 2015). Whether a similar approach could be applied to determine the regulatory networks important for thymus development based on a considerably smaller sample set (consisting of TEPCs and TECs) available in our lab should be explored.

In Chapter 6, it will be important to determine the role, if any, of TGFβ signalling in 3PP and thymus development. This could be done through inhibition or activation of this pathway using either genetic approaches or through the use of chemical inhibitors, as described in this chapter. Analysis of nuclear SMAD2 and/or SMAD3 in the developing thymic primordium could help determine whether this pathway is active during development. Data from biological replicates should be generated to reproducibly determine the effect of TGFβ inhibition on *Foxn1* in adult TECs. An important question is whether the observed effect of TGFβ inhibition on *Foxn1* expression is a direct or indirect. Comparison of the nuclear localization of SMAD2, SMAD3, and SMAD4 between inhibitor treated and control samples could help address this issue. Comparison of results from performing similar experiments using SMAD specific inhibitors and those for other downstream transduction pathways activated by TGFβ receptors could also help address this issue. Finally, it would be interesting to determine whether a simultaneous activation of BMP signalling and an inhibition of TGFβ signalling could further increase the expression of *Foxn1* in TEPCs and TECs.

8. Appendix

8.1 Identification and manipulation of SMAD4 binding sites in *Foxn1* promoter

SMAD4 is a key downstream effector protein of both TGFβ and BMP signalling pathways. While these pathways have been implicated in regulation of *Foxn1*, in the case of BMP signalling in the literature, and in the case of TGFβ signalling through the results shown in Chapter 6, the mechanisms underpinning these effects have not yet been reported. To determine whether the TGFβ and BMP signalling pathways regulate *Foxn1* expression directly through SMAD proteins, I identified SMAD4 binding sites within *Foxn1* promoter as described in Chapter 4. Figure 8.1A shows the position weight matrix (PWM) for SMAD4 in the TRANSFAC database. Several SMAD4 binding sites were identified in the *Foxn1* promoter, as shown in Table 8.1. All of the identified sites had a high degree of match with the SMAD4 PWM suggesting that these sites could be occupied by SMAD4. The locations of these binding sites relative to *Foxn1* TSS and the H3K4me3 marked region is shown in Figure 8.1B and Figure 8.2A. Of these sites, only 2 sites showed perfect match to the SMAD4 PWM (identified by the red box in Figure 8.2A).

The physical binding of SMAD4 on DNA has been investigated in various cell types using ChIP-seq allowing the identification of its target genes. However, given the difficulties in obtaining sufficient number of cells for performing ChIP on transcription factors, this approach was not pursued here. An alternate *in-vivo* approach is to delete or mutate the binding sites for a particular transcription factor in the vicinity of its potential target gene, to determine the functional significance of these sites. The advances in gene editing technologies such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspersed short palindromic repeats (CRISPR)-Cas9 now allows rapid and precise editing of the genome using these techniques. Thus, I decided to use the CRISPR-Cas9 system to test the importance of the selected SMAD4 binding sites in regulation of *Foxn1* expression in TEPCs.

A: SMAD4 position weight matrix

B

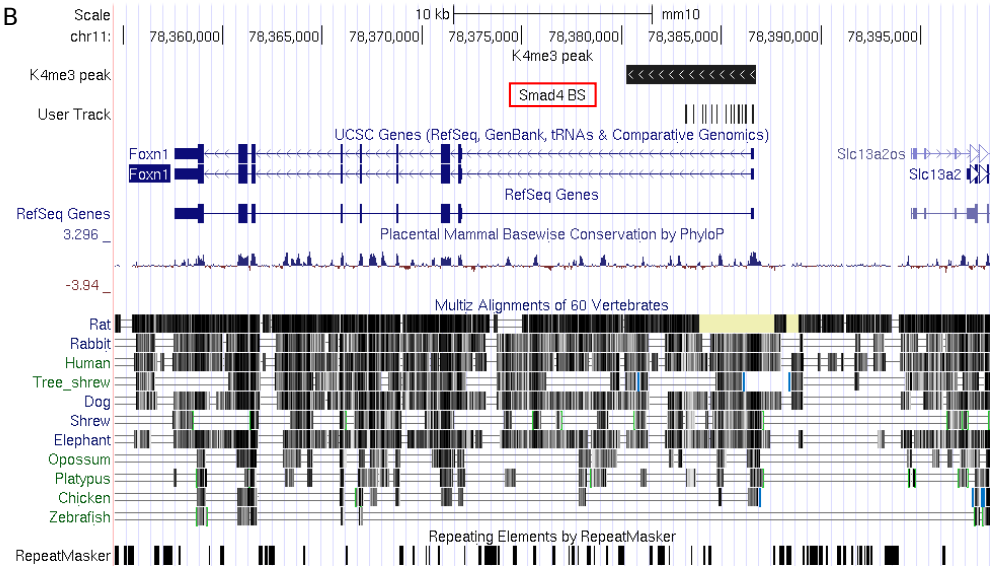


Figure 8.1: Identification of SMAD4 binding sites within *Foxn1* promoter.
The *Foxn1* promoter identified in Chapter 4 was analysed for presence of SMAD4 binding sites using Match algorithm and TRASFAC professional database. (A) SMAD4 position weight matrix in TRANSFAC professional database used to identify binding sites. (B) UCSC genome browser image showing individual SMAD4 binding sites as verticle bars above Foxn1 gene. Also shown is the H3K4me3 marked region and sequence conservation tracks.

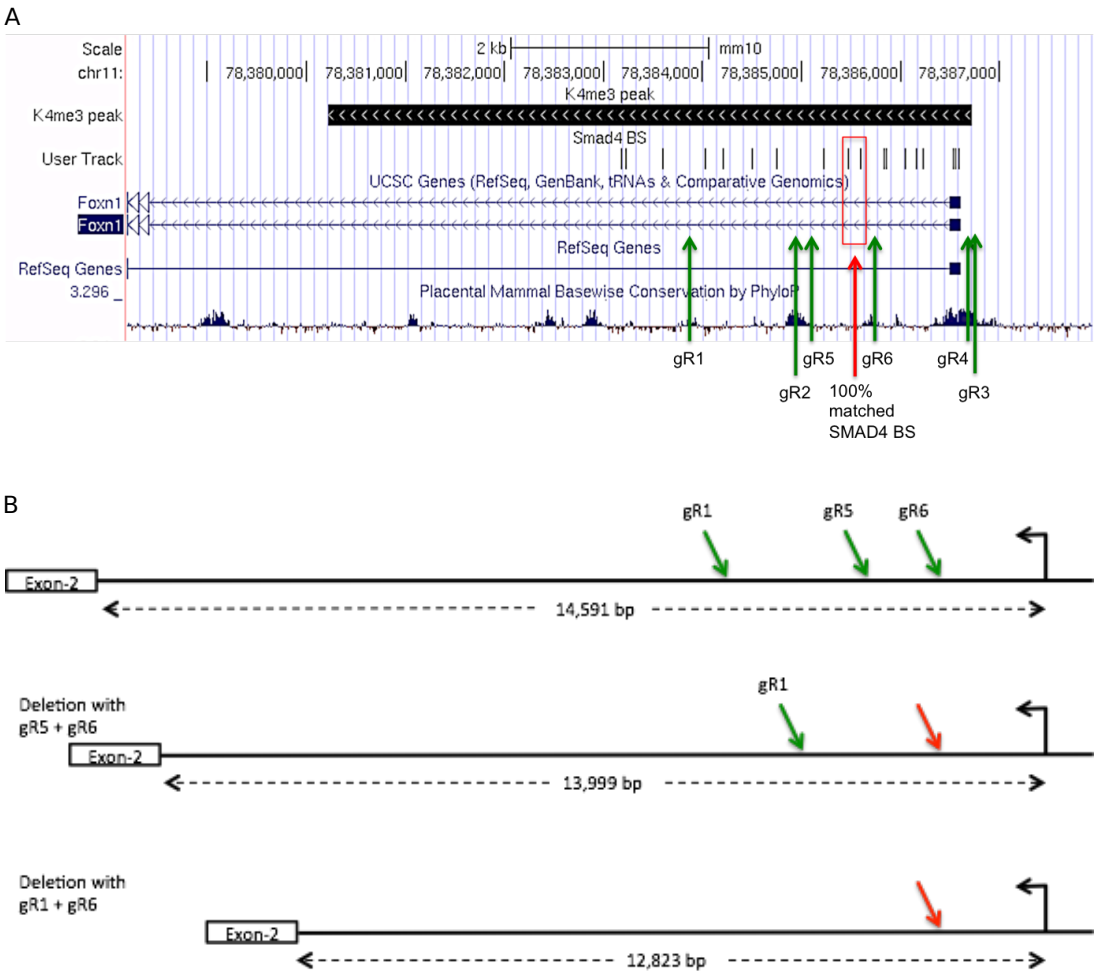


Figure 8.2: Deletion of SMAD4 binding sites within *Foxn1* promoter.
(A) UCSC genome browser view showing the location of individual SMAD4 binding sites within *Foxn1* promoter. The red box shows two SMAD4 sites with 100% match to SMAD4 position weight matrix. Also shown are locations of guide RNAs designed to delete the genomic regions containing SMAD4 binding sites using CRISPR-Cas9. (B) Schematic diagram showing the expected deletions using two guide RNA combinations: gR5 + gR6 and gR1 + gR6. The red arrows show the location of deletion in *Foxn1* locus.

To this end, I designed guide-RNAs (gRNA) spanning the identified SMAD4 binding sites in *Foxn1* promoter using E-CRISP (Heigwer et al. 2014), a web-based application for designing gRNAs. E-CRISP allows identification of target sequences complementary to gRNAs and ending in 3` protospacer-adjacent motif (PAM) and evaluates the off-target effects and target-site homology using Bowtie2 (Heigwer et al. 2014). The *Foxn1* promoter sequence described above was used as the target sequence for E-CRISP, and gRNAs were identified using the default parameters. The identified gRNAs were further filtered for those with the strongest PAM site (NGG) to allow for efficient binding of *S.pyogenes* Cas9. This resulted in 6 gRNAs, which passed all of the filters and were used for subsequent analysis. The sequences of the identified gRNAs and their location relative to *Foxn1* promoter are shown in Table 8.2 and Figure 8.2A, respectively. Among these 6 gRNAs, gR1, gR2, and gR5 were located upstream (on the positive strand) of the two perfectly matched SMAD4 binding sites and gR3, gR4, and gR6 were located downstream. Thus, the SMAD4 binding sites could be deleted using different combinations of upstream and downstream gRNAs, resulting in deletion of genomic regions of varying lengths. This approach was therefore chosen over the alternate approach of mutating the SMAD4 binding sites through homologous recombination for the following reasons. First, the identified promoter region is a part of the 27-kb genomic region shown to be able to recapitulate *Foxn1* expression pattern in the developing thymus, as described in Chapter 1. Thus, deletion of regions within the promoter would allow determination of the relative contributions of these regions in regulating *Foxn1* expression. Secondly, while this approach does not allow determination of the specific role of SMAD4 in regulation of *Foxn1* expression, because the deleted regions will contain binding sites for many other transcription factors, the presence of binding sites for other factors increases the possibility of observing an effect on *Foxn1* expression. The relative contributions of the different transcription factors with binding sites within this region could then be determined by mutating these sites using homologous recombination. This could potentially allow identification of other transcription factors besides *Smad4* that play an important role in regulation *Foxn1* transcription in TEPCs.

Matrix	Start (mm10)	End (mm10)	Strand	Core similarity	Matrix similarity	Sequence
V$SMAD4_Q6_01	78383175	78383182	(-)	1	0.982	tCAGACg
V$SMAD4_Q6_01	78383219	78383226	(-)	1	0.98	cCAGACt
V$SMAD4_Q6_01	78383590	78383597	(+)	1	0.979	gGTCTGa
V$SMAD4_Q6_01	78384020	78384027	(+)	1	0.98	aGTCTGg
V$SMAD4_Q6_01	78384200	78384207	(-)	0.934	0.935	aTAGACa
V$SMAD4_Q6_01	78384491	78384498	(+)	1	0.979	gGTCTGt
V$SMAD4_Q6_01	78384748	78384755	(-)	0.934	0.935	aTAGACa
V$SMAD4_Q6_01	78385215	78385222	(+)	1	0.979	gGTCTGt
V$SMAD4_Q6_01	78385467	78385474	(+)	1	1	tGTCTGt
V$SMAD4_Q6_01	78385594	78385601	(-)	1	1	cCAGACa
V$SMAD4_Q6_01	78385827	78385834	(+)	1	0.979	gGTCTGt
V$SMAD4_Q6_01	78385853	78385860	(+)	1	0.979	gGTCTGc
V$SMAD4_Q6_01	78386042	78386049	(-)	1	0.979	gCAGACc
V$SMAD4_Q6_01	78386154	78386161	(+)	0.934	0.935	tGTCTAt
V$SMAD4_Q6_01	78386225	78386232	(-)	1	0.979	aCAGACc
V$SMAD4_Q6_01	78386529	78386536	(+)	1	0.979	gGTCTGg
V$SMAD4_Q6_01	78386554	78386561	(+)	1	0.982	cGTCTGt
V$SMAD4_Q6_01	78386585	78386592	(+)	1	0.979	gGTCTGg

Table 8.1: SMAD4 binding sites within *Foxn1* promoter.
Shown are all the SMAD4 binding sites identified within the *Foxn1* promoter identified in Chapter 4. The genomic location of the binding sites, core and matrix similarities, and the matched genomic sequences are also shown. Highlighted in red are sites with 100% core and matrix similarity with SMAD4 matrix.

guide RNA	sequence	start (mm10)	end (mm10)	position relative to two perfectly matched Smad4 BS
gR1	ACCCGAGAAGGCAATTTCCT NGG	78383922	78383945	upstream
gR2	CCGGGAGCTGAGCTCGGCAT NGG	78384977	78385000	upstream
gR5	TGGGACCAGGGGAAGAACAG NGG	78385098	78385121	upstream
gR4	ACACAGGTGCATGTCCCAAC NGG	78386928	78386951	downstream
gR3	CTCCGTGAACACACATGTGG NGG	78386966	78386989	downstream
gR6	CTTCTTTGGATGCTGAAGG NGG	78385675	78385695	downstream

Table 8.2: guideRNA sequences and genomic locations.
Shown are the guideRNAs identified using E-crisp software. The sequences, genomic coordinates, and position relative to the two perfectly matched SMAD4 binding sites are also shown.

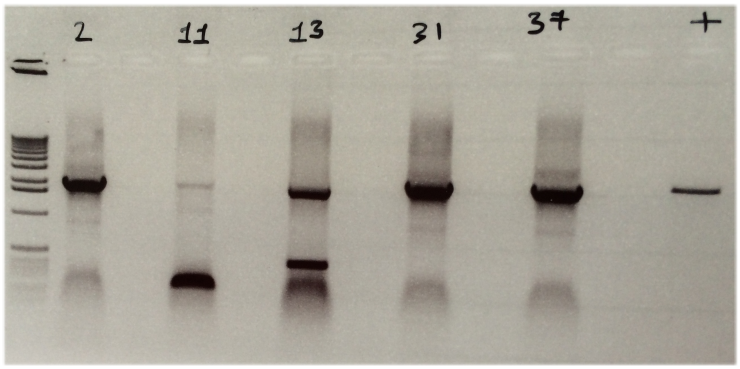
The gRNAs were cloned into pKLV-U6gRNA_BbsI_PGKPuro2ABFP_gRNA vector obtained from Dr Keisuke Kaji's lab (unpublished) (SCRM, University of Edinburgh), which results in gRNA being transcribed using U6 promoter and allows selection of the transfected cells through the expression of puromycin resistance and blue fluorescent protein genes. Plasmids were delivered using lentiviral vectors into a transgenic ES line, also obtained from Dr Keisuke Kaji's lab, which expresses the Cas9 gene under a constitutive promoter (unpublished). The gR1 and gR6, or gR5 and gR6 gRNAs were introduced into ES cells, in order to create deletions of varying lengths within *Foxn1* promoter. The genomic regions deleted using these combinations of gRNAs are shown in Figure 8.2B. The combined transfection with gR1 and gR6 leads to deletion of 1768bp, containing seven putative SMAD4 binding sites. On the other hand, gR5 and gR6 together lead to a deletion of 592bp, containing three putative SMAD4 binding sites.

ES cells transfected with the above combinations of gRNAs were purified by flow cytometry, using the blue fluorescent protein present in the plasmid, 48-hours after infection and were then cultured at clonal density to allow identification of clones containing the desired deletion on both alleles of *Foxn1*. Individual clones were picked and cultured to obtain sufficient cells for subsequent analysis of deleted regions and use in future experiments. As shown in Figure 8.3, of the colonies obtained from ES cells transfected with gR1 and gR6, clone 11 showed deletion of the desired region in both alleles, and clone-13 showed deletion on one allele only. Similarly, of the colonies obtained from ES cells transfected with gR5 and gR6, clone-21 showed deletion on both alleles. Clones carrying homozygous deletions were expanded and frozen in liquid nitrogen for further experiments.

8.1.1 Discussion

SMAD proteins are the primary downstream effector molecules of the TGFβ and the BMP signalling pathways. Thus, it is reasonable to assume that at least some of the observed effects of these pathways in TEPCs and TECs are mediated through SMAD proteins. Given that SMAD4 is the common SMAD protein important in both of

A: genotyping ES clones transfected with gR1 and gR6



B: genotyping ES clones transfected with gR5 and gR6

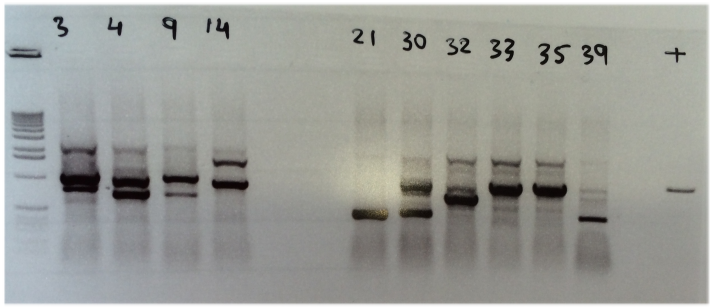


Figure 8.3: Screening of ES cell clones for CRISPR-Cas9 mediated deletion.
Agarose gel electrophoresis images showing deletions within ES cell clones transfected with either (A) gR1 and gR6 or (B) gR5 and gR6 guide RNAs. The first lane show 1kb plus marker used to determine the sizes of observed DNA bands. A PCR product for undeleted region (last lane) was used as positive control for the PCR reaction. Clone 11 in (A) and clones 21 in (B) showed the desired deletions.

these signalling pathways, I focused on analysing the involvement of SMAD4 in regulation of *Foxn1* transcription. The two different cell lines described above will provide useful tools for testing the importance of some of the identified SMAD4 binding sites and also testing the role of deleted genomic region in regulation of *Foxn1* expression *in vivo*. Chimeric embryos should be generated using these ES cell lines to enable analysis of the importance of the deleted regions in development of the thymus and regulation of *Foxn1* expression in TEPCs and TECs. The thymi from such chimeric embryos would be expected to present defects in initiation of *Foxn1* expression, owing to the importance of BMP signalling for this process; and would also be expected to demonstrate an absence of effect of TGFβ signalling on *Foxn1* expression *in vitro* and *in vivo*. Such transgenic mice could also be aged to determine whether this region is important for the age-related decrease in *Foxn1* expression and for TGFβ and/or BMP mediated effects on thymic involution. These transgenic ES cell lines could also be used for direct differentiation into thymic epithelial cells, to test the importance of the deleted regions *in vitro*. The region(s) whose deletion results in changes in *Foxn1* expression could be further tested for binding and regulation by transcription factors, particularly SMAD4, using a combination of *in vivo* and *in vitro* assays, including mutation of the sites using CRISPR-Cas9 mediated homology recombination, and luciferase assays. The binding of SMAD4 to the identified binding sites in the *Foxn1* promoter could be tested by performing ChIP-qPCR using thymic epithelial cell lines, which might overcome the issue of sample availability, can further help distinguish between direct and in-direct effect of TGFβ and BMP signalling pathways. However, the TEC cell lines do not faithfully represent their *in vivo* counterparts due to absence of the 3D structure, other stromal cells, and thymocytes and also differ in the expression of key genes, including *Foxn1*. Thus, these cell lines do not present a suitable model for studying TEC biology and may not be well suited for the above purpose. Note that the TGFβ and BMP signalling pathways can also modulate transcription through SMAD independent mechanisms. The approach described above would not be useful in studying the role of SMAD independent mechanisms in thymus development and function and in regulating *Foxn1* expression. One way to test for this would be through chemical inhibition of SMAD independent pathways in TECs.

8.2 Transcription factor binding sites in *Dll4*, *Ccl25*, *p63*, *Kitl*, and *Gata3* promoters identified using MATCH and TRANSFAC

Dll4 promoter:

AIRE, AML3, AP-1, AP-2alphaA, Arid5a, BBX secondary motif, BEN, Blimp-1, BRCA1:USF2, Cdc5, CDP CR1, Cdx-2, C/EBPalpha, Churchill, CIZ, COE1, CP2, CPBP, CREB1, ATF-2, CRX, CTCF, DEAF1, deltaEF1, dlx-3, LXR, PXR, CAR, COUP, RAR, DRI1, E2A, E2F, Ebox, Egr-1, ER-alpha, Ets, FAC1, ZNF263, Freac-3, GATA, GCMb, GEN_INI, Gfi1, GKLF, GLI, Hbp1, HES-1, Hic1, HIF-1alpha, HMGIY, HNF-1alpha, HNF-3beta, HSF1, Ikaros, ING4, INSM1, ipf1, IRF-1, islet1, Kaiso, LEF-1, LRH-1, LUN-1, MAF, MAFA, MAZ, MAZR, MECP2, MEF-2, MEIS1, MRF2, MTF-1, c-Myb, myogenin, MZF-1, NF-1, NF-1A, NF-AT1, NF-Y, Nkx2.5, p53, Pbx, Pit-1, POU2F1, POU6F1, RBP-Jkappa, RelA-p65, REST, RFX, RFX1, RNF96, RREB-1, RUSH-1alpha, SF-1, Smad4, Sox10, Sp1, SP100 seondary motif, SREBP, SRY, STAT1, TATA, Tbx5, TEF-1, TTF-1, XBP-1, Xvent-1, YY1, ZFP105 secondary motif, ZFP161, Zfx, ZNF333, ZSCAN4 secondary motif

Ccl25 promoter:

AHR, AIRE, AML3, AP-1, AP-2alphaA, Arid5a, BBX secondary motif, Bcl-6, BEN, Blimp-1, BRCA1:USF2, Cdc5, CDP CR1, Cdx-2, C/EBPalpha, Churchill, CIZ, COE1, CP2, CPBP, ATF-2, CRX, CTCF, deltaEF1, DMRT4, LXR, PXR, CAR, COUP, RAR, DRI1, E2A, E2F, Ebox, EKLF, ER-alpha, Ets, FAC1, ZNF263, Freac-3, GATA, GCMb, GEN_INI, Gfi1, GKLF, GLI, HES-1, Hic1, HIF-1alpha, HMGIY, HNF-1alpha, HNF-3beta, HNF-4A, HNF-6, HSF1, Ikaros, ING4, INSM1, ipf1, IRF-1, islet1, Kaiso, LEF-1, LRH-1, MAF, MAFA, MAZ, MECP2, MEF-2, mef-2A, MEIS1, MRF2, MTF-1, c-Myb, myogenin, NF-1, NF-1A, NF-AT1, NF-AT5, NF-Y, Nkx2.5, p53, Pbx, Pit-1, POU2F1, POU6F1, RBP-Jkappa, RelA-p65, REST, RFX, RFX1, RNF96, RORalpha, RREB-1, RUSH-1alpha, SF-1, Smad4, Sox10, Sox2, Sp1, SP100 seondary motif, SREBP, SRY, STAT1, TATA, Tbx5, TEF-1, TTF-1, Xvent-1, YY1, ZFP105 secondary motif, ZFP161, Zfx, ZNF333, ZSCAN4 secondary motif

p63 promoter:

AIRE, AML3, AP-1, AP-2alphaA, Arid5a, BBX secondary motif, Bcl-6, BEN, Blimp-1, BRCA1:USF2, Cdc5, CDP CR1, Cdx-2, C/EBPalpha, Churchill, CP2, CPBP, CREB1, ATF-

2, CRX, CTCF, deltaEF1, DMRT4, LXR, PXR, CAR, COUP, RAR, DRI1, E2A, E2F, Ebox, Egr-1, ER-alpha, Ets, FAC1, ZNF263, Freac-3, GATA, GCMb, GEN_INI, Gfi1, GKLF, GLI, Hbp1, HES-1, Hic1, HIF-1alpha, HMGIY, HNF-1alpha, HNF-3beta, HNF-6, HSF1, Ikaros, ING4, INSM1, ipf1, IRF-1, islet1, Kaiso, LEF-1, LRH-1, LUN-1, MAF, MAFA, MAZ, MAZR, MEF-2, mef-2A, MEIS1, MRF2, c-Myb, myogenin, NF-1, NF-1A, NF-AT1, NF-Y, Nkx2.5, p53, Pbx, Pit-1, POU2F1, POU6F1, RBP-Jkappa, RelA-p65, RFX1, RREB-1, RUSH-1alpha, SF-1, Smad4, Sox10, Sox2, Sp1, SP100 seondary motif, SREBP, SRY, STAT1, TATA, Tbx5, TEF-1, TTF-1, Xvent-1, YY1, ZFP105 secondary motif, ZFP161, ZNF333, ZSCAN4 secondary motif

***Kitl* promoter:**

AHR, AIRE, AP-1, AP-2alphaA, Bbx, BBX secondary motif, BEN, BRCA1:USF2, CDP CR1, Cdx-2, C/EBPalpha, Churchill, CPBP, CREB1, ATF-2, CRX, CTCF, deltaEF1, DRI1, E2A, E2F, Ebox, Egr-1, ER-alpha, Ets, FAC1, ZNF263, Freac-3, GATA, GEN_INI, Gfi1, GKLF, GLI, HES-1, Hic1, HIF-1alpha, HMGIY, HNF-1alpha, HNF-3beta, HSF1, Ikaros, ING4, ipf1, islet1, Kaiso, LEF-1, LRH-1, MAF, MAFA, MAZ, MECP2, MEF-2, mef-2A, MEIS1, MTF-1, c-Myb, myogenin, NF-1, NF-1A, NF-AT1, Nkx2.5, p53, Pbx, Pit-1, POU2F1, POU6F1, RBP-Jkappa, RelA-p65, RFX, RFX1, RNF96, RREB-1, RUSH-1alpha, SF-1, Smad4, Sox10, Sox2, Sp1, SP100 seondary motif, SRY, TATA, Tbx5, TEF-1, TTF-1, Xvent-1, YY1, ZFP105 secondary motif, ZFP161, ZNF333, ZSCAN4 secondary motif

***Gata3* promoter:**

AHR, AIRE, AML3, AP-1, AP-2alphaA, Bbx, BBX secondary motif, BEN, Blimp-1, BRCA1:USF2, CDP CR1, Cdx-2, C/EBPalpha, Churchill, CIZ, COE1, CP2, CPBP, CREB1, ATF-2, CRX, CTCF, deltaEF1, DMRT4, LXR, PXR, CAR, COUP, RAR, DRI1, E2A, E2F, Ebox, Egr-1, ER-alpha, Ets, FAC1, ZNF263, Freac-3, GATA, GCMb, GEN_INI, Gfi1, GKLF, GLI, Hbp1, HES-1, Hic1, HIF-1alpha, HMGIY, HNF-1alpha, HNF-3beta, HNF-4A, HNF-6, Ikaros, ING4, ipf1, IRF-1, islet1, Kaiso, LEF-1, LRH-1, MAF, MAFA, MAZ, MAZR, MEF-2, mef-2A, MRF2, MTF-1, c-Myb, myogenin, MZF-1, NF-1, NF-1A, NF-AT1, NF-AT5, NF-Y, Nkx2.5, p53, Pbx, Pit-1, POU2F1, POU6F1, RBP-Jkappa, RelA-p65, REST, RFX, RFX1, RNF96, RREB-1, RUSH-1alpha, SF-1, Smad4, Sox10, Sox2, Sp1, SP100 seondary motif, SREBP, SRY, STAT1, TATA, Tbx5, TEF-1, TTF-1, Xvent-1, YY1, ZFP105 secondary motif, ZFP161, Zfx, ZNF333, ZSCAN4 secondary motif

List of Figures

List of Tables

References

Abu-Issa, R. et al., 2002. Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. *Development*, 129(19), pp.4613–4625.

Adli, M. & Bernstein, B., 2011. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nature protocols*, 6(10), pp.1656–68.

Adli, M., Zhu, J. & Bernstein, B.E., 2010. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nature methods*, 7(8), pp.615–8.

Akiyama, T. et al., 2008. The Tumor Necrosis Factor Family Receptors RANK and CD40 Cooperatively Establish the Thymic Medullary Microenvironment and Self-Tolerance. *Immunity*, 29(3), pp.423–437.

Allman, D. et al., 2003. Thymopoiesis independent of common lymphoid progenitors. *Nature immunology*, 4(2), pp.168–174.

Anders, S. & Huber, W., 2010. Differential expression analysis for sequence count data. *Genome biology*, 11(10), p.R106.

Andersson, E.R., Sandberg, R. & Lendahl, U., 2011. Notch signaling: simplicity in design, versatility in function. *Development*, 138(17), pp.3593–3612.

Bae, S. et al., 2000. The bHLH gene Hes6 , an inhibitor of Hes1 , promotes neuronal differentiation. *Development*, 127, pp.2933–2943.

Bajoghli, B. et al., 2009. Evolution of genetic networks underlying the emergence of thymopoiesis in vertebrates. *Cell*, 138(1), pp.186–97.

Balciunaite, G. et al., 2002. Wnt glycoproteins regulate the expression of FoxN1, the gene defective in nude mice. *Nat Immunol*, 3(11), pp.1102–1108.

Balciunaite, G., Ceredig, R. & Rolihk, A.G., 2005. The earliest subpopulation of mouse thymocytes contains potent T, significant macrophage, and natural killer cell but no B-lymphocyte potential. *Blood*, 105(5), pp.1930–1936.

Balwierz, P.J. et al., 2014. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Research*, 24(5), pp.869–884.

Benayoun, B. a., Caburet, S. & Veitia, R. a., 2011. Forkhead transcription factors: Key players in health and disease. *Trends in Genetics*, 27(6), pp.224–232.

Bennett, A.R. et al., 2002. Identification and Characterization of Thymic Epithelial Progenitor Cells. *Immunity*, 16, pp.803–814.

Birnbaum, R.Y. et al., 2012. *Coding exons function as tissue-specific enhancers of nearby genes.*,

Blackburn, C.C. et al., 1996. The nu gene acts cell-autonomously and is required for differentiation of thymic epithelial progenitors. *Proc. Natl. Acad. Sci. USA*, 93(12), pp.5742–5746.

Blackburn, C.C. & Manley, N.R., 2004. Developing a new paradigm for thymus organogenesis. *Nature reviews. Immunology*, 4(4), pp.278–289.

Bleul, C.C. et al., 2006. Formation of a functional thymus initiated by a postnatal epithelial progenitor cell. *Nature*, 441(7096), pp.992–6.

Bleul, C.C. & Boehm, T., 2005. BMP Signaling Is Required for Normal Thymus Development. *Journal of immunology*, 175, pp.5213–5221.

Bleul, C.C. & Boehm, T., 2000. Chemokines define distinct microenvironments in the developing thymus. *European Journal of Immunology*, 30, pp.3371–3379.

Boehm, T. & Bleul, C.C., 2007. The evolutionary history of lymphoid organs. *Nature immunology*, 8(2), pp.131–135.

Bossard, P. & Zaret, K.S., 1998. GATA transcription factors as potentiators of gut endoderm differentiation. *Development*, 125(24), pp.4909–4917.

Boyd, R.L. et al., 1993. Inside the thymus The thvmic microenvironment. *Immunology today*, 14, pp.445–59.

Bredenkamp, N., Ulyanchenko, S., et al., 2014. An organized and functional thymus generated from FOXN1-reprogrammed fibroblasts. *Nature Cell Biology*, 1(August), pp.1–7.

Bredenkamp, N., Nowell, C.S. & Blackburn, C.C., 2014. Regeneration of the aged thymus by a single transcription factor. *Development (Cambridge, England)*, 141(8), pp.1627–37.

Bryson, J.L. et al., 2013. Cell-autonomous defects in thymic epithelial cells disrupt endothelial-perivascular cell interactions in the mouse thymus. *PloS one*, 8(6), p.e65196.

Buenrostro, J.D. et al., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin , DNA-binding proteins and nucleosome position. *Nature methods*, 10(12), pp.1213–1218.

Cai, J. et al., 2009. Genetic interplays between Msx2 and Foxn1 are required for Notch1 expression and hair shaft differentiation. *Developmental biology*, 326(2), pp.420–430.

Calderón, L. & Boehm, T., 2012. Synergistic, context-dependent, and hierarchical functions of epithelial components in thymic microenvironments. *Cell*, 149(1), pp.159–72.

Calo, E. & Wysocka, J., 2013. Modification of enhancer chromatin: what, how, and why? *Molecular cell*, 49(5), pp.825–37.

Carlsson, P. & Mahlapuu, M., 2002. Forkhead Transcription Factors: Key Players in Development and Metabolism. *Developmental Biology*, 250, pp.1–23.

Chapman, D.L. et al., 1996. Expression of the T-box Family Genes, Tbxl-Tbx5, During Early Mouse Development. *Developmental dynamics*, 206, pp.379–390.

Chen, L., Xiao, S. & Manley, N.R., 2009. Foxn1 is required to maintain the postnatal thymic microenvironment in a dosage-sensitive manner. *Blood*, 113(3), pp.567–74.

Chen, Y. et al., 2012. Systematic evaluation of factors influencing ChIP-seq fidelity. *Nature methods*, 9(6), pp.609–614.

Cheng, L. et al., 2010. Postnatal tissue-specific disruption of transcription factor FoxN1 triggers acute thymic atrophy. *The Journal of biological chemistry*, 285(8), pp.5836–47.

Chojnowski, J.L. et al., 2014. Multiple roles for HOXA3 in regulating thymus and parathyroid differentiation and morphogenesis in mouse. *Development*, 141(19), pp.3697–3708.

Cirillo, L. a. et al., 1998. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO Journal*, 17(1), pp.244–254.

Cirillo, L.A. et al., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, 9(2), pp.279–289.

Clark, K.L. et al., 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, 363, pp.412–420.

Coffer, P.J. & Burgering, B.M.T., 2004. Forkhead-box transcription factors and their role in the immune system. *Nature reviews. Immunology*, 4(11), pp.889–99.

Cordier, A.C. & Haumont, S.M., 1980. Development of thymus, parathyroids, and ultimo-branchial bodies in NMRI and nude mice. *The American journal of anatomy*, 157(3), pp.227–263.

Cordier, A.C. & Heremans, J.F., 1975. Nude mouse embryo: ectodermal nature of the primordial thymic defect. *Scand J Immunol*, 4(2), pp.193–196.

Core Team, R., 2015. R: A language and environment for statistical computing. Available at: http://www.r-project.org/.

Crabtree, G.R. & Olson, E.N., 2002. NFAT signaling: Choreographing the social lives of cells. *Cell*, 109, pp.67–79.

Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA*, 107(50), pp.21931–21936.

Criswell, T.L. & Arteaga, C.L., 2007. Modulation of NFkappaB activity and E-cadherin by the type III transforming growth factor beta receptor regulates cell growth and motility. *Journal of Biological Chemistry*, 282(44), pp.32491–32500.

Cunliffe, V.T., Furley, A.J.W. & Keenan, D., 2002. Complete rescue of the nude mutant phenotype by a wild-type Foxn1 transgene. *Mammalian genome : official journal of the International Mammalian Genome Society*, 13(5), pp.245–52.

Depreter, M.G.L. et al., 2007. Identification of Plet-1 as a specific marker of early thymic epithelial progenitor cells. , (19), pp.18–23.

Dooley, J. et al., 2007. FGFR2IIIb signaling regulates thymic epithelial differentiation. *Developmental dynamics*, 236(12), pp.3459–3471.

Le Douarin, N.M. & Jotereau, F. V, 1975. Tracing of cells of the avian thymus through embryonic life in interspecific chimeras. *Journal of Experimental Medicine*, 142, pp.17–40.

Ellis, S.L. et al., 2013. The role of Tenascin C in the lymphoid progenitor cell niche. *Experimental Hematology*, 41(12), pp.1050–1061.

Erickson, M. et al., 2002. Regulation of thymic epithelium by keratinocyte growth factor. *Blood*, 100(9), pp.3269–3278.

Ernst, J. et al., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345), pp.43–9.

Felli, M.P. et al., 1999. Expression pattern of Notch1, 2 and 3 and Jagged1 and 2 in lymphoid and stromal thymus components: Distinct ligand-receptor interactions in intrathymic T cell development. *International Immunology*, 11(7), pp.1017–1025.

Feng, J. et al., 2012. Identifying ChIP-seq enrichment using MACS. *Nature protocols*, 7(9), pp.1728–40.

Flanagan, S.P., 1966. “Nude” a new hairless gene with pleiotropic effects in the mouse. *Genet Res*, 8, pp.295–309.

Foster, K. et al., 2008. Contribution of Neural Crest-Derived Cells in the Embryonic and Adult Thymus. *The Journal of Immunology*.

Foster, K.E. et al., 2010. EphB-ephrin-B2 interactions are required for thymus migration during organogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30), pp.13414–13419.

Frank, D.U. et al., 2002. An Fgf8 mouse mutant phenocopies human 22q11 deletion syndrome. *Development*, 129(19), pp.4591–4603.

Fujimoto, Y. et al., 2002. CD83 expression influences CD4+ T cell development in the thymus. *Cell*, 108(6), pp.755–767.

Gajiwala, K.S. & Burley, S.K., 2000. Winged helix proteins. *Current Opinion in Structural Biology*, 10(1), pp.110–116.

Gardiner, J.R. et al., 2012. Localised inhibition of FGF signalling in the third pharyngeal pouch is required for normal thymus and parathyroid organogenesis. *Development*, 139(18), pp.3456–66.

Garfin, P.M. et al., 2013. Inactivation of the RB family prevents thymus involution and promotes thymic function by direct control of Foxn1 expression. *The Journal of experimental medicine*, 210(6), pp.1087–97.

Garg, V. et al., 2001. Tbx1, a DiGeorge Syndrom Candidate Gene, Is Regulated by Sonic Hedgehog during Phayrngeal Arch Development. *Developmental biology*, 235, pp.62–73.

Gill, J. et al., 2002. Generation of a complete thymic microenvironment by MTS24(+) thymic epithelial cells. *Nature immunology*, 3(7), pp.635–42.

Godfrey, D.I. et al., 1993. A developmental pathway involving four phenotypically and functionally distinct subsets of CD3-CD4-CD8- triple-negative adult mouse thymocytes defined by CD44 and CD25 expression. *Journal of immunology (Baltimore, Md. : 1950)*, 150(10), pp.4244–4252.

Godfrey, D.I., Zlotnik, a & Suda, T., 1992. Phenotypic and functional characterization of c-kit expression during intrathymic T cell development. *Journal of immunology (Baltimore, Md. : 1950)*, 149(7), pp.2281–2285.

Gommeaux, J. et al., 2009. Thymus-specific serine protease regulates positive selection of a subset of CD4+ thymocytes. *European Journal of Immunology*, 39(4), pp.956–964.

Gordon, J. et al., 2010. Evidence for an early role for BMP4 signaling in thymus and parathyroid morphogenesis. *Developmental biology*, 339(1), pp.141–154.

Gordon, J. et al., 2004. Functional evidence for a single endodermal origin for the thymic epithelium. *Nature immunology*, 5(5), pp.546–53.

Gordon, J. et al., 2001. Gcm2 and Foxn1 mark early parathyroid- and thymus-specific domains in the developing third pharyngeal pouch. *Mechanisms of development*, 103(1-2), pp.141–3.

Griffith, A. V et al., 2009. Increased thymus- and decreased parathyroid-fated organ domains in Splotch mutant embryos. *Developmental biology*, 327(1), pp.216–227.

Grigorieva, I. V et al., 2010. Gata3-deficient mice develop parathyroid abnormalities due to dysregulation of the parathyroid-specific transcription factor Gcm2. *J Clin Invest.*, 120(6).

Gualdi, R. et al., 1996. Hepatic specification of the gut endoderm in vitro: Cell signaling and transcriptional control. *Genes and Development*, 10(13), pp.1670–1682.

Guo, C. et al., 2011. A Tbx1-Six1/Eya1-Fgf8 genetic pathway controls mammalian cardiovascular and craniofacial morphogenesis. *The Journal of Clinical Investigation*, 121(4), pp.1585–1595.

Halbritter, F., Vaidya, H.J. & Tomlinson, S.R., 2011. GeneProf: analysis of high-throughput sequencing experiments. *Nature Methods*, 9(1), pp.7–8.

Hannenhalli, S. & Kaestner, K.H., 2009. The evolution of Fox genes and their role in development and disease. *Nature reviews. Genetics*, 10(4), pp.233–240.

Harada, H. et al., 1994. Structure and regulation of the human interferon regulatory factor 1 (IRF-1) and IRF-2 genes: implications for a gene network in the interferon system. *Molecular and cellular biology*, 14(2), pp.1500–1509.

Hauri-Hohl, M. et al., 2014. A regulatory role for TGF-β signaling in the establishment and function of the thymic medulla. *Nature immunology*, 15(6), pp.554–61.

Hauri-Hohl, M.M. et al., 2008. TGF-beta signaling in thymic epithelial cells regulates thymic involution and postirradiation reconstitution. *Blood*, 112(3), pp.626–34.

He, W. et al., 2002. Overexpression of Smad7 results in severe pathological alterations in multiple epithelial tissues. *The EMBO journal*, 21(11), pp.2580–2590.

Heigwer, F., Kerr, G. & Boutros, M., 2014. E-CRISP: fast CRISPR target site identification. *Nature methods*, 11(2), pp.122–3.

Heintzman, N.D. et al., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, 39(3), pp.311–8.

Hendriks, R.W. et al., 1999. Expression of the transcription factor GATA-3 is required for the development of the earliest T cell progenitors and correlates with stages of cellular proliferation in the thymus. *European Journal of Immunology*, 29(6), pp.1912–1918.

Hetzer-Egger, C. et al., 2002. Thymopoiesis requires Pax9 function in thymic epithelial cells. *European Journal of Immunology*, 32(4), pp.1175–1181.

Hikosaka, Y. et al., 2008. The Cytokine RANKL Produced by Positively Selected Thymocytes Fosters Medullary Thymic Epithelial Cells that Express Autoimmune Regulator. *Immunity*, 29(3), pp.438–450.

Hogan, P.G. et al., 2003. Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes & development*, 17(18), pp.2205–2232.

Huang, D.W., Sherman, B.T. & Lempicki, R. a., 2009. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1), pp.1–13.

Huang, D.W., Sherman, B.T. & Lempicki, R.A., 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols*, 4(1), pp.44–57.

Huh, S.-H. & Ornitz, D.M., 2010. Beta-catenin deficiency causes DiGeorge syndrome-like phenotypes through regulation of Tbx1. *Development*, 137(7), pp.1137–47.

Itoi, M., Tsukamoto, N., Yoshida, H., et al., 2007. Mesenchymal cells are required for functional development of thymic epithelial cells. *International immunology*, 19(8), pp.953–64.

Itoi, M. et al., 2001. Two distinct steps of immigration of hematopoietic progenitors into the early thymus anlage. *International Immunology*, 13(9), pp.1203–1211.

Itoi, M., Tsukamoto, N. & Amagai, T., 2007. Expression of Dll4 and CCL25 in Foxn1-negative epithelial cells in the post-natal thymus. *International immunology*, 19(2), pp.127–32.

Jeker, L.T. et al., 2008. Maintenance of a normal thymic microenvironment and T-cell homeostasis require Smad4-mediated signaling in thymic epithelial cells. *Blood*, 112(9), pp.3688–95.

Jenkinson, W.E., Jenkinson, E.J. & Anderson, G., 2003. Differential requirement for mesenchyme in the proliferation and maturation of thymic epithelial progenitors. *The Journal of experimental medicine*, 198(2), pp.325–332.

Jerome, L. a & Papaioannou, V.E., 2001. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nature genetics*, 27(3), pp.286–91.

Jiang, X. et al., 2000. Fate of the mammalian cardiac neural crest. *Development*, 127(8), pp.1607–1616.

Jin, X. et al., 2014. Long-Term Persistence of Functional Thymic Epithelial Progenitor Cells In Vivo under Conditions of Low FOXN1 Expression. *PLoS ONE*, 9(12), p.e114842.

Kaestner, K.H., Knochel, W. & Martinez, D.E., 2000. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes & development*, 14(2), pp.142–146.

Kagey, M.H. et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314), pp.430–435.

Kalatzis, V. et al., 1998. Eya Expression in the Developing Ear and Kidney: Towards the Understanding of the Pathogenesis of Branchio-Oto-Renal (BOR) Syndrome. *Developmental dynamics*, 499, pp.486–499.

Kato, S., 1997. Thymic microvascular system. *Microscopy Research and Technique*, 38(3), pp.287–299.

Kel, A. et al., 2006. Beyond microarrays: find key transcription factors controlling signal transduction pathways. *BMC bioinformatics*, 7 Suppl 2, p.S13.

Kel, A.E. et al., 2003. MATCH^{TM}: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*, 31(13), pp.3576–3579.

Ki, S. et al., 2014. Global Transcriptional Profiling Reveals Distinct Functions of Thymic Stromal Subsets and Age-Related Changes during Thymic Involution. *Cell Reports*, 9(1), pp.402–415.

Klug, D.B. et al., 2002. Cutting edge: thymocyte-independent and thymocyte-dependent phases of epithelial patterning in the fetal thymus. *Journal of immunology (Baltimore, Md. : 1950)*, 169(6), pp.2842–2845.

Koch, U. et al., 2008. Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *The Journal of experimental medicine*, 205(11), pp.2515–2523.

Kolde, R., 2015. pheatmap: Pretty Heatmaps. Available at: http://cran.r-project.org/package=pheatmap.

Koyano-Nakagawa, N. et al., 2000. Hes6 acts in a positive feedback loop with the neurogenins to promote neuronal differentiation. *Development*, 127(19), pp.4203–16.

Krumlauf, R., 1994. Hox genes in vertebrate development. *Cell*, 78(2), pp.191–201.

Kumar, R., Langer, J.C. & Snoeck, H., 2006. Transforming growth factor- beta2 is involved in quantitative genetic variation in thymic involution. *Blood*, 107(5), pp.1974–1979.

Kurooka, H. et al., 1996. Rescue of the hairless phenotype in nude mice by transgenic insertion of the wild-type Hfh11 genomic locus. *International immunology*, 8(6), pp.961–966.

Lai, K.-P. et al., 2013. Targeting thymic epithelia AR enhances T-cell reconstitution and bone marrow transplant grafting efficacy. *Molecular endocrinology (Baltimore, Md.)*, 27(1), pp.25–37.

Landt, S.G. et al., 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research*, 22, pp.1813–1831.

Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), p.R25.

Lee, C.S. et al., 2005. The initiation of liver development is dependent on Foxa transcription factors. *Nature*, 435, pp.944–947.

Lelli, K.M., Slattery, M. & Mann, R.S., 2012. Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. *Annual review of genetics*, 46, pp.43–68.

Lenhard, B., Sandelin, A. & Carninci, P., 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature reviews. Genetics*, 13(4), pp.233–45.

Leung, D.W. et al., 1989. Vascular endothelial growth factor is a secreted angiogenic mitogen. *Science*, 246(4935), pp.1306–1309.

Liberzon, A. et al., 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27(12), pp.1739–1740.

Lin, C.-C. et al., 2014. Bhlhe40 controls cytokine production by T cells and is essential for pathogenicity in autoimmune neuroinflammation. *Nature communications*, 5, pp.3551–3564.

Lind, E.F. et al., 2001. Mapping precursor movement through the postnatal thymus reveals specific microenvironments supporting defined stages of early lymphoid development. *Journal of Experimental Medicine*, 194(2), pp.127–134.

Lindahl, P. et al., 1997. Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science*, 277(5323), pp.242–245.

Lindsay, E. a, 2001. Chromosomal microdeletions: dissecting del22q11 syndrome. *Nature reviews. Genetics*, 2(11), pp.858–868.

Lindsay, E. a et al., 1999. Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature*, 401(6751), pp.379–383.

Liu, C. et al., 2006. Coordination between CCR7- and CCR9-mediated chemokine signals in prevascular fetal thymus colonization. *Blood*, 108(8), pp.2531–2539.

Liu, C. et al., 2005. The role of CCL21 in recruitment of T-precursor cells to fetal thymi. *Blood*, 105(1), pp.31–39.

Liu, X. & Bosselut, R., 2004. Duration of TCR signaling controls CD4-CD8 lineage differentiation in vivo. *Nature immunology*, 5(3), pp.280–288.

Liu, Y. et al., 2008. A critical function for TGF-beta signaling in the development of natural CD4+CD25+Foxp3+ regulatory T cells. *Nature immunology*, 9(6), pp.632–640.

Lmmerts van Beuren, K. et al., 2010. Hes1 expression is reduced in Tbx1 null cells and is required for the development of structures affected in 22q11 deletion syndrome. *Developmental Biology*, 340(2), pp.369–380.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *bioRxiv*, pp.1–21.

Luc, S. et al., 2012. The earliest thymic T cell progenitors sustain B cell and myeloid lineage potential. *Nature immunology*, 13(4), pp.412–9.

Luo, W. et al., 2009. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics*, 10, p.161.

Luo, W. & Brouwer, C., 2013. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 29(14), pp.1830–1831.

Ma, D. et al., 2012. Foxn1 maintains thymic epithelial cells to support T-cell development via mcm2 in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America*, 109(51), pp.21040–5.

Macatee, T.L. et al., 2003. Ablation of Specific Expression Domains Reveals Discrete Functions of Ectoderm- and Endoderm-Derived FGF8 During Cardiovascular and Pharyngeal Development. *Development*, 130(25), pp.6361–6374.

Manley, N.R., 2000. Thymus organogenesis and molecular mechanisms of thymic epithelial cell differentiation. *Seminars in immunology*, 12(5), pp.421–428.

Manley, N.R. & Capecchi, M.R., 1998. Hox group 3 paralogs regulate the development and migration of the thymus, thyroid, and parathyroid glands. *Developmental Biology*, 195(1), pp.1–15.

Manley, N.R. & Capecchi, M.R., 1995. The role of Hoxa-3 in mouse thymus and thyroid development. *Development*, 121, pp.1989–2003.

Marski, M. et al., 2005. CD18 is required for optimal development and function of CD4+CD25+ T regulatory cells. *Journal of immunology (Baltimore, Md. : 1950)*, 175(12), pp.7889–7897.

Massagué, J., 2012. TGFβ signalling in context. *Nature Molecular Cell Biology*, 13, pp.616–630.

Masuda, K. et al., 2009. Notch activation in thymic epithelial cells induces development of thymic microenvironments. *Molecular Immunology*, 46(8-9), pp.1756–1767.

Matsuzaki, Y. et al., 1993. Characterization of c-kit positive intrathymic stem cells that are restricted to lymphoid differentiation. *The Journal of experimental medicine*, 178(4), pp.1283–1292.

Matys, V. et al., 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic acids research*, 34(Database issue), pp.D108–D110.

McLean, C.Y. et al., 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology*, 28(5), pp.495–501.

Mecklenburg, L. et al., 2001. The nude mouse skin phenotype: the role of Foxn1 in hair follicle development and cycling. *Experimental and molecular pathology*, 71(2), pp.171–178.

Meier, N., Dear, T.N. & Boehm, T., 1999. Whn and mHa3 are components of the genetic hierarchy controlling hair follicle differentiation. *Mechanisms of Development*, 89(1-2), pp.215–221.

Moignard, V. et al., 2013. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nature Cell Biology*, 15(4), pp.363–372.

Moignard, V. et al., 2015. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nature Biotechnology*, 33(3), pp.269–280.

Mootha, V.K. et al., 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*, 34(3), pp.267–273.

Morgan, M. et al., 2009. ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, 25(19), pp.2607–2608.

Mori, K. et al., 2007. The perivascular space as a path of hematopoietic progenitor cells and mature T cells between the blood circulation and the thymic parenchyma. *International Immunology*, 19(6), pp.745–753.

Mullen, A.C. et al., 2011. Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell*, 147(3), pp.565–76.

Müller, S.M. et al., 2005. Gene targeting of VEGF-A in thymus epithelium disrupts thymus blood vessel architecture. *Proc. Natl. Acad. Sci. USA*, 102(30), pp.10587–10592.

Müller, S.M. et al., 2008. Neural Crest Origin of Perivascular Mesenchyme in the Adult Thymus. *The Journal of Immunology*, 180, pp.5344–5351.

Nagle, D.L. et al., 1994. Structural analysis of chromosomal rearrangements associated with the developmental mutations Ph , W19H , and Rw on mouse chromosome 5. *Proc. Natl. Acad. Sci. USA*, 91(July), pp.7237–7241.

Nehls, M. et al., 1994. New member of the winged-helix protein family disrupted in mouse and rat nude mutations. *Nature*, 372, pp.103–107.

Nehls, M. et al., 1996. Two genetically separable steps in the differentiation of thymic epithelium. *Science (New York, N.Y.)*, 272(5263), pp.886–889.

Neubüser, A., Koseki, H. & Balling, R., 1995. Characterization and Developmental Expression of Pax9, a Paired-Box-Containing Gene Related to Pax1. *Developmental biology*, (170), pp.701–716.

Neves, H. et al., 2012. Modulation of Bmp4 signalling in the epithelial-mesenchymal interactions that take place in early thymus and parathyroid development in avian embryos. *Developmental biology*, 361(2), pp.208–19.

Nowell, C.S. et al., 2011. Foxn1 regulates lineage progression in cortical and medullary thymic epithelial cells but is dispensable for medullary sublineage divergence. *PLoS genetics*, 7(11), p.e1002348.

De Obaldia, M.E., Bell, J.J. & Bhandoola, A., 2013. Early T-cell progenitors are the major granulocyte precursors in the adult mouse thymus. *Blood*, 121(1), pp.64–71.

Obsil, T. & Obsilova, V., 2008. Structure/function relationships underlying regulation of FOXO transcription factors. *Oncogene*, 27(16), pp.2263–2275.

Odaka, C. et al., 2013. TGF-β type II receptor expression in thymic epithelial cells inhibits the development of Hassall's corpuscles in mice. *International immunology*, 25(11), pp.633–42.

Okubo, T. et al., 2011. Ripply3, a Tbx1 repressor, is required for development of the pharyngeal apparatus and its derivatives in mice. *Development*, 138(2), pp.339–48.

Ortman, C.L., 2002. Molecular characterization of the mouse involuted thymus: aberrations in expression of transcription regulators in thymocyte and epithelial compartments. *International Immunology*, 14(7), pp.813–822.

Ouyang, W. et al., 2013. TGF-β cytokine signaling promotes CD8+ T cell development and low-affinity CD4+ T cell homeostasis by regulation of interleukin-7 receptor α expression. *Immunity*, 39(2), pp.335–46.

Ouyang, W. et al., 2010. Transforming growth factor-beta signaling curbs thymic negative selection promoting regulatory T cell development. *Immunity*, 32(5), pp.642–53.

Owen, J.J. & Ritter, M., 1969. Tissue Interaction In The Development of Thymus Lymphocytes. *Journal of Experimental Medicine*, 129, pp.431–442.

Paek, H. et al., 2011. β-Catenin-dependent FGF signaling sustains cell survival in the anterior embryonic head by countering Smad4. *Developmental cell*, 20(5), pp.689–99.

Pantelouris, E.M., 1968. Absence of thymus in a mouse mutant. *Nature*, 217, pp.370–371.

Parent, A. V et al., 2013. Generation of Functional Thymic Epithelium from Human Embryonic Stem Cells that Supports Host T Cell Development. *Cell stem cell*, 13(2), pp.219–29.

Patel, S.R. et al., 2006. Bmp4 and Noggin expression during early thymus and parathyroid organogenesis. *Gene expression patterns : GEP*, 6(8), pp.794–9.

Peters, H. et al., 1999. Pax1 and Pax9 synergistically regulate vertebral column development. *Development*, 126, pp.5399–5408.

Peters, H. et al., 1998. Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes & Development*, 12(17), pp.2735–2747.

Picelli, S. et al., 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols*, 9(1), pp.171–81.

Porritt, H.E. et al., 2004. Heterogeneity among DN1 prothymocytes reveals multiple progenitors with different capacities to generate T cell and non-T cell lineages. *Immunity*, 20(6), pp.735–745.

Potter, C.S. et al., 2011. The nude mutant gene Foxn1 is a HOXC13 regulatory target during hair follicle and nail differentiation. *The Journal of investigative dermatology*, 131(4), pp.828–837.

Rada-Iglesias, A. et al., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), pp.279–83.

Ramsköld, D. et al., 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology*, 5(12), pp.1–11.

Rau, A. et al., 2013. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics*, 29(17), pp.2146–2152.

Reeh, K.A.G. et al., 2014. Ectopic TBX1 suppresses thymic epithelial cell differentiation and proliferation during thymus organogenesis. *Development*, 141, pp.2950–2958.

Revest, J.M. et al., 2001. Development of the thymus requires signaling through the fibroblast growth factor receptor R2-IIIb. *Journal of immunology (Baltimore, Md. : 1950)*, 167(4), pp.1954–1961.

Rossi, S.W. et al., 2006. Clonal analysis reveals a common progenitor for thymic cortical and medullary epithelium. *Nature*, 441(7096), pp.988–991.

Rossi, S.W. et al., 2007. RANK signals from CD4(+)3(-) inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. *The Journal of experimental medicine*, 204(6), pp.1267–1272.

Sansom, S.N. et al., 2014. Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome research*, 12, pp.1918–1931.

Schlake, T., 2005. FGF signals specifically regulate the structure of hair shaft medulla via IGF-binding protein 5. *Development*, 132(13), pp.2981–90.

Schlake, T. & Boehm, T., 2001. Expression domains in the skin of genes affected by the nude mutation and identified by gene expression profiling. *Mechanisms of Development*, 109(2), pp.419–422.

Sempowski, G.D. et al., 2000. Leukemia inhibitory factor, oncostatin M, IL-6, and stem cell factor mRNA expression in human thymus increases with age and is associated with thymic atrophy. *Journal of immunology (Baltimore, Md. : 1950)*, 164(4), pp.2180–2187.

Seoane, J. et al., 2004. Integration of Smad and Forkhead Pathways in Growth Control and Cancer. *Cell*, 117, pp.211–223.

Shakib, S. et al., 2009. Checkpoints in the development of thymic cortical epithelial cells. *Journal of immunology (Baltimore, Md. : 1950)*, 182(1), pp.130–137.

Shen, Y. et al., 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature*, 488(7409), pp.116–20.

Su, D. et al., 2003. A domain of Foxn1 required for crosstalk-dependent thymic epithelial cell differentiation. *Nature immunology*, 4(11), pp.1128–35.

Su, D. et al., 2001. Hoxa3 and Pax1 regulate epithelial cell death and proliferation during thymus and parathyroid organogenesis. *Developmental biology*, 236(2), pp.316–29.

Su, D. & Manley, N.R., 2000. Hoxa3 and Pax1 Transcription Factors Regulate the Ability of Fetal Thymic Epithelial Cells to Promote Thymocyte Development. *The Journal of Immunology*, 164, pp.5753–5760.

Subramanian, A. et al., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, 102(43), pp.15545–15550.

Sultan, M. et al., 2008. A Global View of Gene Activity and Alternative Splicing by Deep Sequencing of the Human Transcriptome. , 685(August), pp.956–961.

Swann, J.B. et al., 2014. Conversion of the Thymus into a Bipotent Lymphoid Organ by Replacement of Foxn1 with Its Paralog , Foxn4. *Cell reports*, 8(4), pp.1184–1197.

Takahama, Y. et al., 1994. Early Progression of Thymocytes along the CD4/CD8 Developmental Pathway Is Regulated by a Subset of Thymic Epithelial Cells Expressing Transforming Growth Factor beta. *The Journal of experimental medicine*, 179(May), pp.1495–1506.

Takahama, Y., 2006. Journey through the thymus: stromal guides for T-cell development and selection. *Nature reviews. Immunology*, 6(2), pp.127–135.

Theiler, K., 1989. *The House Mouse: atlas of Embryonic Development*, Springer New York.

Ting, C.-N. et al., 1996. Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature*, 384, pp.474–478.

Trapnell, C., Pachter, L. & Salzberg, S.L., 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), pp.1105–1111.

Tsai, K.L. et al., 2006. Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins. *Journal of Biological Chemistry*, 281(25), pp.17400–17409.

Tsukamoto, N. et al., 2005. Lack of Delta like 1 and 4 expressions in nude thymus anlages. *Cellular Immunology*, 234, pp.77–80.

Vitelli, F. et al., 2002. Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways. *Human molecular genetics*, 11(8), pp.915–922.

Viviiie, S. et al., 1993. Mice Lacking the MHC Class II-Associated Invariant Chain. *Cell*, 72, pp.635–648.

Wallin, J. et al., 1996. Pax1 is expressed during development of the thymus epithelium and is required for normal T-cell maturation. *Development*, 122, pp.23–30.

Wang, J. et al., 2006. Defective ALK5 signaling in the neural crest leads to increased postmigratory neural crest cell apoptosis and severe outflow tract defects. *BMC developmental biology*, 6, p.51.

Wei, Q. & Condie, B.G., 2011. A focused in-situ hybridization screen identifies candidate transcriptional regulators of thymic epithelial cell development and function. *PloS one*, 6(11), p.e26795.

Weigel, D. et al., 1989. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. *Cell*, 57(4), pp.645–658.

Winnier, G. et al., 1995. Bone morphogenetic protein-4 Is reqmrea for mesoderm formation and patterning in the mouse. *Genes & Development*, pp.2105–2116.

Wurbel, M.A. et al., 2001. Mice lacking the CCR9 CC-chemokine receptor show a mild impairment of early T- and B-cell development and a reduction in T-cell receptor γδ+ gut intraepithelial lymphocytes. *Blood*, 98(9), pp.2626–2632.

Wurdak, H. et al., 2005. Inactivation of TGF β signaling in neural crest stem cells leads to multiple defects reminiscent of DiGeorge syndrome. *Genes and Development*, 19, pp.530–535.

Xu, H., Cerrato, F. & Baldini, A., 2005. Timed mutation and cell-fate mapping reveal reiterated roles of Tbx1 during embryogenesis , and a crucial function during segmentation of the pharyngeal system via regulation of endoderm expansion. *Development and disease*, 132(19), pp.4387–4395.

Xu, P. et al., 2002. Eya1 is required for the morphogenesis of mammalian thymus , parathyroid and thyroid. *Development*, 129, pp.3033–3044.

Xu, P.X. et al., 1997. Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. *Development*, 124(1), pp.219–231.

Zaret, K.S. & Carroll, J.S., 2011. Pioneer transcription factors: Establishing competence for gene expression. *Genes and Development*, 25(21), pp.2227–2241.

Zhang, Y. et al., 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology*, 9(9), p.R137.

Zhang, Z. & Wu, W.-S., 2013. Sodium butyrate promotes generation of human induced pluripotent stem cells through induction of the miR302/367 cluster. *Stem cells and development*, 22(16), pp.2268–77.

Zook, E.C. et al., 2013. Enhancing T lineage production in aged mice: a novel function of Foxn1 in the bone marrow niche. *Journal of immunology (Baltimore, Md. : 1950)*, 191(11), pp.5583–93.

Zook, E.C. et al., 2011. Overexpression of Foxn1 attenuates age-associated thymic involution and prevents the expansion of peripheral CD4 memory T cells. *Blood*, 118, pp.5723–5731.

Zou, D. et al., 2008. Eya1 gene dosage critically affects the development of sensory epithelia in the mammalian inner ear. *Human Molecular Genetics*, 17(21), pp.3340–3356.

Zou, D. et al., 2006. Patterning of the third pharyngeal pouch into thymus / parathyroid by Six and Eya1. *Developmental Biology*, 293, pp.499–512.